VOLUME 29 · NUMBER 1 · JANUARY 1 2011

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Gene Expression Signature to Improve Prognosis Prediction of Stage II and III Colorectal Cancer

Ramon Salazar, Paul Roepman, Gabriel Capella, Victor Moreno, Iris Simon, Christa Dreezen, Adriana Lopez-Doriga, Cristina Santos, Corrie Marijnen, Johan Westerga, Sjoerd Bruin, David Kerr, Peter Kuppen, Cornelis van de Velde, Hans Morreau, Loes Van Velthuysen, Annuska M. Glas, Laura J. Van't Veer, and Rob Tollenaar

From the Institut Català d'Oncologia–Biomedical Research Institute of Bellvitge, L'Hospitalet de Llobregat; University of Barcelona, Barcelona, Spain; Agendia; Netherlands Cancer Institute; Slotervaart Hospital, Amsterdam; Leiden University Medical Center, Leiden, the Netherlands; and University of Oxford, Radcliffe Infirmary, Oxford, United Kingdom.

Submitted May 5, 2010; accepted September 7, 2010; published online ahead of print at www.jco.org on November 22, 2010.

RNA isolation and hybridization of the samples and part of the analysis were performed and funded at Agendia. The training set of the study was partly supported by the Leiden Medical Centre Institutional Grant and by the Dutch Genomics Initiative Cancer Genomics Center in the Netherlands Cancer Institute. The validation of the study was partly supported by the Catalan Institute of Oncology and the Private Foundation of the Biomedical Research Institute of Bellvitge, the Spanish Ministry of Science (Grants No. SAF 06-6084 and SAF 2009-07319), the Instituto de Salud Carlos III (Grants No. PI08-1635 and PI09-01037), Spanish Networks Red Temática de Investigación Cooperativa en Cáncer (Grant No. RD06/0020/1050), Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública G55 and the Accion Transversal del Cancer, the Catalan Government Departament d'Universitats, Recerca i Societat de la Informació (Grants No. 2009SGR1489 and 2009SGR290), the European Commission (Grant No. FP7-COOP-Health-2007-B), HiperDart, and Fundació Gastroenterologia Dr Francisco Vilardell.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Ramon Salazar, MD, Institut Català d'Oncologia-IDIBELL, L'Hospitalet de Llobregat, Av Gran Via 199-203, Barcelona, Spain 08907; e-mail: ramonsalazar@iconcologia.net.

0732-183X/11/2901-17/$20.00

DOI: 10.1200/JCO.2010.30.1077

ABSTRACT

Purpose
This study aims to develop a robust gene expression classifier that can predict disease relapse in patients with early-stage colorectal cancer (CRC).

Patients and Methods
Fresh frozen tumor tissue from 188 patients with stage I to IV CRC undergoing surgery was analyzed using Agilent 44K oligonucleotide arrays. Median follow-up time was 65.1 months, and the majority of patients (83.6%) did not receive adjuvant chemotherapy. A nearest mean classifier was developed using a cross-validation procedure to score all genes for their association with 5-year distant metastasis–free survival.

Results
An optimal set of 18 genes was identified and used to construct a prognostic classifier (ColoPrint). The signature was validated on an independent set of 206 samples from patients with stage I, II, and III CRC. The signature classified 60% of patients as low risk and 40% as high risk. Five-year relapse-free survival rates were 87.6% (95% CI, 81.5% to 93.7%) and 67.2% (95% CI, 55.4% to 79.0%) for low- and high-risk patients, respectively, with a hazard ratio (HR) of 2.5 (95% CI, 1.33 to 4.73; $P = .005$). In multivariate analysis, the signature remained one of the most significant prognostic factors, with an HR of 2.69 (95% CI, 1.41 to 5.14; $P = .003$). In patients with stage II CRC, the signature had an HR of 3.34 ($P = .017$) and was superior to American Society of Clinical Oncology criteria in assessing the risk of cancer recurrence without prescreening for microsatellite instability (MSI).

Conclusion
ColoPrint significantly improves the prognostic accuracy of pathologic factors and MSI in patients with stage II and III CRC and facilitates the identification of patients with stage II disease who may be safely managed without chemotherapy.

J Clin Oncol 29:17-24. © 2010 by American Society of Clinical Oncology

INTRODUCTION

The American Joint Committee on Cancer TNM staging system is the current standard for determining the prognosis of patients with colorectal cancer (CRC). Patients with stage I CRC have a 5-year survival rate of approximately 93%, which decreases to approximately 80% for patients with stage II disease and to 60% for patients with stage III disease.[1] Despite numerous clinical trials, the benefit of adjuvant chemotherapy for patients with stage II CRC is still debatable.[2-4] In Western countries, official guidelines give suggestions for risk stratification but no clear recommendations on the administration of adjuvant chemotherapy.[5] In contrast, adjuvant treatment is universally recommended for all patients with stage III disease.[6] However, patients with T1-2N1M0 tumors (stage IIIA) have significantly higher survival rates than patients with stage IIB tumors,[1] suggesting that adjuvant chemotherapy selection needs optimization.

To date, substantial effort has been put into the identification of clinicopathologic parameters that predict prognosis of patients with stage II disease. The most important factors for predicting the risk of systemic recurrence (ie, distant metastases) are emergency presentation, poorly differentiated tumor, depth of tumor invasion, and adjacent organ involvement (T4).[5,7] Inadequate sampling of lymph nodes is an additional risk factor.[8] Among the molecular factors investigated as prognostic candidates in early CRCs, microsatellite instability (MSI) is the

only one that has remained significant both in a meta-analysis and prospective trials.[9-11]

During the last decade, gene expression profiling has shown great promise in predicting the long-term outcome of an individual patient.[12] The power of applying customized microarray technology to predict the prognosis of patients with breast cancer has led to the successful development of a US Food and Drug Administration–approved breast cancer prognostic test (MammaPrint; Agendia, Amsterdam, the Netherlands).[13,14] Several studies have already described prognostic gene expression profiles for patients with CRC from tumor samples[10,15-21] and even from adjacent normal mucosa.[22] However, few studies compared the genomic prognosis prediction with traditional risk factors except for stage.[20] In this study, we demonstrate the development and validation of a new prognosis signature to distinguish low- and high-risk patients using gene expression analysis.

PATIENTS AND METHODS

Patients and Tumor Samples

Samples used for classifier training (n = 188) were prospectively collected between 1983 and 2002 at the Netherlands Cancer Institute (Amsterdam), the Leiden University Medical Center (Leiden), and the Slotervaart General Hospital (Amsterdam) in the Netherlands. Samples for the validation set of patients (n = 206) were prospectively collected at the Institut Catala d'Oncologia in Barcelona, Spain, between 1996 and 2004. Clinical and pathologic data were extracted from the medical records and centrally reviewed for the purpose of this study. Patients with rectal cancer underwent total mesorectal excision controlled surgery. Patients were staged according to the American Joint Committee on Cancer TNM staging system and monitored for relapse (development of distant metastases or locoregional recurrence) and overall survival (median follow-up time: training set, 65.1 months; validation set, 54.8 months). Detailed patient information is listed in Table 1 and Appendix Table A1 (online only). The study was approved by the medical ethical boards of the participating medical centers. In all, 71.6% of the patients did not receive adjuvant chemotherapy (83.6% and 61.9% of patients in the training and validation sets, respectively).

Table 1. Patient Demographics and Clinical Characteristics for the Training and Validation Sets

Demographic or Clinical Characteristic	Training Set (n = 188)		Validation Set (n = 206)	
	No. of Patients	%	No. of Patients	%
Hospital				
LUMC	76	40.4		
NKI	52	27.7		
Slotervaart	48	25.5		
Other	12	6.4		
ICO Barcelona			206	100.0
Median age, years	67.9		69	
Median follow-up, months	65.1		54.1	
Sex				
Male	84	44.7	132	64.1
Female	104	55.3	74	35.9
Localization				
Left	92	50.3	115	55.8
Right	74	40.4	67	32.5
Rectum	17	9.3	24	11.7
Stage				
I (T2 only)	24	12.8	30	14.6
II	100	53.2	114	55.3
III	56	29.8	62	30.1
IV	8	4.2	—	—
Grade				
1	11	5.8	90	43.7
2	141	75.0	100	48.5
3	30	16.0	16	7.8
NA	6	3.2	—	—
Distant metastasis				
No	137	72.9	173	84.0
Yes	51	27.1	33	16.0
Chemotherapy				
No	148	78.7	125	60.7
Yes	36	19.1	77	37.4
Unknown	4	2.1	4	1.9

Abbreviations: LUMC, Leiden University Medical Center; NKI, Netherlands Cancer Institute; ICO, Institut Catala d'Oncologia; NA, not available.

Mutational and MSI Analysis

Mutations in *BRAF* V600; *KRAS* codons 12, 13, and 61; and *PI3KCA* exons 9 and 20 were assessed in cDNA by means of direct sequencing of polymerase chain reaction products using M13 primers after reverse transcriptase polymerase chain reaction. Primers used and experimental conditions are available on request. In the training set, 5-μm slides were immunohistochemically stained for the markers MLH1 and PMS2 using standard protocols to identify MSI-high (MSI-H) patients. In the validation set, the MSI status analysis was performed as previously described.[23]

Gene Expression Analysis

RNA isolation, labeling, and hybridization to Agendia customized whole-genome oligonucleotide high-density microarrays followed procedures as previously described.[14] Samples were hybridized against a colon cancer reference pool, consisting of primary tumor tissue from 44 patients with CRC. Raw fluorescence intensities were quantified and normalized using Agilent Feature Extraction software (Agilent, Santa Clara, CA) according to the manufacturer's protocols and imported into R/Bioconductor (http://www.bioconductor.org/) for further analysis.

A supervised training approach was performed to identify a prognostic CRC gene signature. Using a cross-validation procedure, all 33,834 gene probes that showed variation across the 188 training samples were scored for their association (*t* test) with 5-year distant metastasis–free survival (DMFS). During each of the leave-one-out cross-validation iterations, the set of genes with a significant DMFS association [abs(T) > 3.5] was marked. From the comprehensive pool of genes, an optimal set of 18 nonredundant probes showed robust DMFS association in more than 50% of all iterations, a selection criterion suggested by Michiels et al.[24] These 18 probes corresponded to 18 unique genes (Appendix Table A2, online only) and were used to construct a nearest centroid–based classifier (called ColoPrint). This type of classifier has been proven to be useful for clinical use[14] and scores a sample as either low risk or high risk for development of distant metastasis. The optimal threshold for the classifier index score was selected to reach optimal sensitivity and specificity in the training set. If a sample's index exceeded the set threshold, it was classified as a high-risk sample; if the index of the sample was below the threshold, the sample was classified as a low-risk sample.

Statistical Analysis

For the analysis of the training set, we defined the probability that patients remain free of distant metastases as the first event. For the analysis of the validation set, the primary end point was relapse-free survival (RFS), which was defined as the probability that patients remain free of recurrence (locoregional or metastatic) as the first event; data on all other patients were censored on the date of the last follow-up visit or date of death. Deaths of no specific cause were censored to evaluate true prognostic prediction. Data were analyzed from the date of surgery to the time of the first event or the date on which data were censored, according to the Kaplan-Meier method, and the curves were compared with use of the log-rank test. To increase the number of events, the training set and validation set were combined to analyze the prognostic

information of *KRAS*, *BRAF*, and *PI3KCA* mutation in univariate analysis. MSI status was analyzed in a subset of patients from the training set (n = 90) and in all patients from the validation set. To determine the independence of our classifier to clinicopathologic variables in predicting an individual's risk of experiencing relapse, we analyzed the validation set using univariate analysis followed by multivariate analysis. Sex; localization of the tumor; T stage; N stage; number of lymph nodes assessed; histologic grade; lymphatic, vascular, and perineural invasion; adjuvant chemotherapy administration; MSI status; and gene expression profile were included as variables in this analysis. Log-rank tests were used in the univariate analysis, and a multivariate Cox model was built including the significant variables from univariate analysis on all patients in the validation set and on subsets of patients with stage II and III disease only. In addition, multivariate analysis with American Society of Clinical Oncology (ASCO) clinical risk criteria (defined as T stage of 4, poor grade tumor, < 13 assessed lymph nodes, or emergency presentation with obstruction or perforation[5]) and ColoPrint as independent variables was performed in patients with stage II disease. All calculations were performed with SPSS statistical package version 16.0 (SPSS, Chicago, IL).

RESULTS

Profile Development

Unsupervised hierarchical clustering of the 188 tumor tissues of the training set revealed three main molecular subtypes. Patients with subtype A showed a good outcome, whereas patients with subtype C had a relative poor outcome (84% and 58% 5-year disease-free survival, respectively; hazard ratio [HR], 1.8; $P = .015$). Most patients (110 of 188 patients) fell into the intermediate prognosis cluster, subtype B (Fig 1). Further investigation of these subtypes indicated that both survival-associated subtypes, A and C, were enriched for patients with an activating *BRAF* V600E mutation. In the subtype A group, 52% of patients had *BRAF* mutations, and in the poor-outcome subtype C group, 22% of patients had *BRAF* mutations; whereas in the subtype B group, 4% of patients were mutation carriers. Subtype A was enriched for patients with MSI (MSI-H). Fifteen of 90 patients in the training set with known MSI status were MSI-H. Thirteen of these 15 patients belonged to subtype A. The molecular subtypes had no correlation with stage (data not shown).

Only subtype B (n = 110) was used to develop a prognostic signature to avoid building a classifier that was mainly based on the extreme expression patterns of subtypes A and C. An optimal set of 18 genes was identified and used to construct the ColoPrint prognostic classifier (see Patients and Methods). The classifier was applied to all samples in the training set (n = 188) using a leave-one-out cross-validation procedure (Appendix Fig A1, online only). Five-year DMFS rates were 82% (95% CI, 76% to 89%) and 50% (95% CI, 38% to 66%) for patients with a low-risk and high-risk signature, respectively (HR, 3.41; 95% CI, 1.95 to 5.91; $P < .001$). Among the 29 patients who received adjuvant chemotherapy, 14 (48.3%) were classified as ColoPrint low risk, and 15 (51.7%) were classified as high risk.

Independent Validation

An independent patient cohort of 206 patients was used to evaluate the performance of the colorectal prognosis classifier (Table 2). Most patients in the validation set were patients with stage II disease (n = 114). In the validation set of all patients, 60% of patients were identified as low risk, whereas 40% of patients were high risk (Fig 2). Low-risk patients had a 5-year RFS rate of 87.6% (95% CI, 81.5% to 93.7%), whereas high-risk patients had a 5-year RFS rate of only 67.2% (95% CI, 55.4% to 79%). Male patients and patients with colon cancer on the left side were more often classified as high risk than female patients and patients with cancer in the right colon. High risk was also positively associated with relapse, time to relapse, and time to death. Classification as ColoPrint low or high risk was not associated with grade, age, stage, or number of assessed lymph nodes (Table 2).

When the classifier was applied to patients with stage II and stage III disease separately, it correctly classified low- and high-risk patients in both groups (Fig 2). In the analysis of patients with stage II disease, 63.2% were classified as low risk, and 36.8% were classified as high risk, with 5-year RFS rates of 90.9% (95% CI, 84% to 97.8%) and 73.9% (95% CI, 59.2% to 88.6%), respectively ($P = .017$).

Comparison to Clinical Factors, Mutational Analysis, and MSI

In the combined training and validation set, in patients with known mutation status (n = 381), *KRAS* mutations were detected in 115 patients (30.2%), *PI3KCA* mutations were detected in 45 patients (11.8%), and *BRAF* mutations were detected in 42 patients (11%). In our data set, the mutations, either alone or in combination, were not predictive for relapse or overall survival (data not shown). MSI status was known for 276 patients (90 patients in the training set, 186 patients in the validation set), of whom 29 were classified as MSI-H. Patients

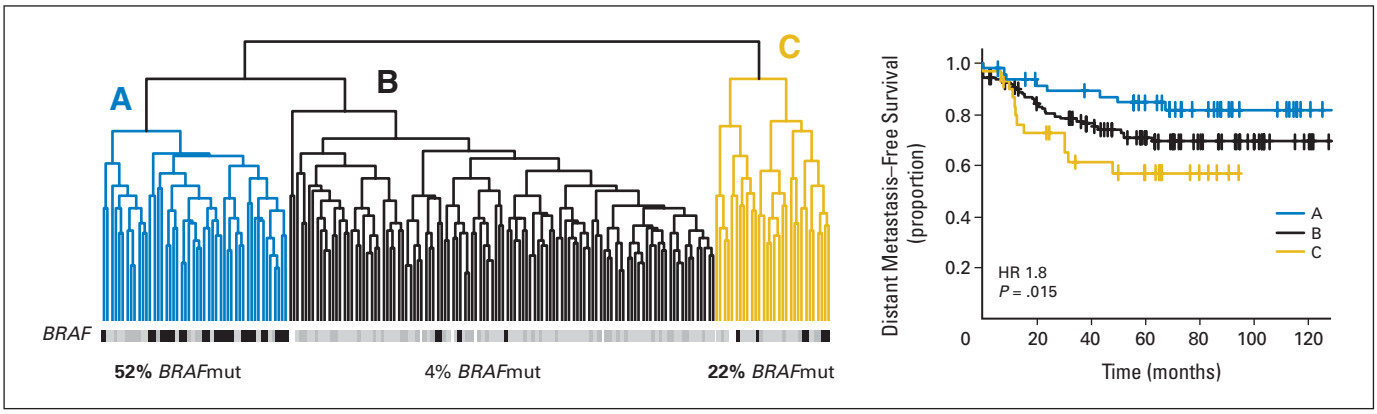

Fig 1. Unsupervised clustering of 188 tumor samples (training set) revealed three molecular subtypes (A, B, and C). *BRAF* mutation status was known for 179 patients and is associated with the subtypes.

Table 2. Association of Clinicopathologic Variables With Assessment of Colon Low- and High-Risk Signature in the Validation Set

Variable	Total (N = 206)		ColoPrint				P
			Low Risk (n = 123)		High Risk (n = 83)		
	No. of Patients	%	No. of Patients	%	No. of Patients	%	
Age, years							
Median	69		69		69.45		.676
≤ 70	113	54.9	68	55.3	45	54.2	.880
> 70	93	45.1	55	44.7	38	45.8	
Localization							.000
Left	115	55.8	57	46.3	58	69.9	
Right	67	32.5	54	43.9	13	15.7	
Rectum	24	11.7	12	9.8	12	14.5	
Grade							.386
1	90	43.7	51	41.5	39	47.0	
2	100	48.5	60	48.8	40	48.2	
3	16	7.8	12	9.8	4	4.8	
Sex							.085
Male	132	64.1	73	59.3	59	71.1	
Female	74	35.9	50	40.7	24	28.9	
No. of LNs assessed							
Median	17.5		19.0		15.0		.093
≤ 12	55	26.7	29	23.6	26	31.3	.218
> 12	151	73.3	94	76.4	57	68.7	
Stage							.405
I	30	14.6	15	12.2	15	18.1	
II	114	55.3	72	58.5	42	50.6	
III	62	30.1	36	29.3	26	31.3	
pT							.471
2	33	16.0	17	13.8	16	19.3	
3	157	76.2	95	77.2	62	74.7	
4	16	7.8	11	8.9	5	6.0	
pN							.334
0	144	69.9	87	70.7	57	68.7	
1	42	20.4	27	22.0	15	18.1	
2	20	9.7	9	7.3	11	13.3	
DM							.027
No	173	84.0	109	88.6	64	77.1	
Yes	33	16.0	14	11.4	19	22.9	
Median time to DM, months	51.8		60		38.7		.001
Relapse (local, regional, distant)							.013
No	166	80.6	106	86.2	60	72.3	
Yes	40	19.4	17	13.8	23	27.7	
Median time to relapse, months	51.8		60		38.7		.001
Death							.846
No	175	85.0	104	84.6	71	85.5	
Yes	31	15.0	19	15.4	12	14.5	
Median survival time, months	54.1		60.2		49.5		.029
Chemotherapy							.971
No	125	61.9	75	62	50	61.7	
Yes	77	38.1	46	38	31	38.3	
MSI-high							.039
No	172	83.5	99	80.0	73	88.0	
Yes	14	6.8	12	9.9	2	2.4	
NA	20	9.7	12	9.9	8	9.6	
Lymphatic invasion							.846
No	175	85.0	104	84.6	71	85.5	
Yes	31	15.0	19	15.4	12	14.5	
Venous invasion							.718
No	192	93.2	114	92.7	78	94.0	
Yes	14	6.8	9	7.3	5	6.0	

(continued on following page)

Table 2. Association of Clinicopathologic Variables With Assessment of Colon Low- and High-Risk Signature in the Validation Set (continued)

Variable	Total (N = 206)		ColoPrint				
			Low Risk (n = 123)		High Risk (n = 83)		
	No. of Patients	%	No. of Patients	%	No. of Patients	%	*P*
Perineural invasion							.566
No	203	98.5	122	99.2	81	97.6	
Yes	3	1.5	1	0.8	2	2.4	
Lymphatic, venous, or perineural invasion							.925
No	162	78.6	97	78.9	65	78.3	
Yes	44	21.4	26	21.1	18	21.7	

NOTE. No patient had obstruction/perforation.
Abbreviations: LN, lymph node; DM, distant metastasis; MSI, microsatellite instability; NA, not available.

with MSI-H were mainly patients with stage II disease (21 of 29 patients; 72%) and had a high frequency of *BRAF* mutation (15 of 29 patients; 52%). These patients were also mostly classified as ColoPrint low risk (26 of 29 patients; 90%), indicating that the good prognosis of the MSI-H patients is identified by the gene classifier. This is also verified in analysis of the validation set only (Table 2).

For comparison of performance of ColoPrint and clinical factors, only results from the validation set were used (Table 2). ColoPrint was the strongest predictor of RFS in the univariate analysis, where only stage, T stage, and lymph node status showed a similar magnitude of statistical significance. Among the 77 patients who received adjuvant chemotherapy, 46 (59.7%) were classified as ColoPrint low risk, and 31 (40.3%) were classified as ColoPrint high risk; chemotherapy administration was not a significant prognostic factor for RFS in this series (Table 3 and Appendix Table A3, online only). In the multivariate analysis of all samples and of samples from patients with stage III disease only, ColoPrint remained a strong independent prognostic factor (Table 4 and Appendix Table A4, online only). Analysis of DMFS yielded similar results (Appendix Tables A5 and A6, online only).

In the subset of patients with stage II disease (n = 114), ColoPrint was the strongest predictor for RFS in the univariate analysis (HR, 3.34; 95% CI, 1.24 to 9.00; $P = .017$) and multivariate analysis (Tables 3 and 4). The analysis of the relative performance of the gene classifier with conventional clinicopathologic factors revealed that T stage was also associated with prognosis (HR, 3.15; 95% CI, 1.02 to 9.69; $P = .045$). In addition, the classifier performed independently from the ASCO risk criteria when they were analyzed either individually (Table 3) or combined (HR, 3.66; 95% CI, 1.24 to 9.08; $P = .017$; Appendix Table A7, online only). Interestingly, a high degree of discordance (48.2%) in risk stratification between ColoPrint and ASCO criteria was observed (Appendix Table A8, online only). Finally, in the subgroup of patients with stage II disease, among 40 patients (36%) who received adjuvant chemotherapy, 28 patients (68%) were classified as ColoPrint low risk, and 12 patients (32%) were classified as high risk, and chemotherapy administration was not a significant prognostic factor for RFS ($P = .34$) or overall survival.

Additional In Silico Validation and Functional Analysis

To further explore the clinical and biologic relevance of ColoPrint, an additional in silico validation and functional analysis of the set of genes included was performed. Gene expression data of 322 stage I to III colorectal tumor samples from three previously published studies[18,20,25] were available for in silico validation of the gene classifier. In the first data set of 100 patients (Gene Expression Omnibus accession GSE5206),[25] ColoPrint risk scores were significantly associated with development of disease recurrence (Wilcoxon $P = .0092$), with an area under the receiver operating curve of 0.68 (data not shown). This data set was combined with two additional data sets (GSE10402 and ArrayExpress accession MEXP-1245),[18,20] yielding a total of 322 colorectal tumors. ColoPrint risk outcome was significantly associated with RFS ($P < .001$, McNemar test), with an odds ratio of 2.8 (95% CI, 1.6 to 4.7). In the analysis of all stages, ColoPrint low-risk samples (n = 177, 55%) showed a 5-year RFS rate of 83.8% (95% CI, 79.3% to 87.5%) compared with a 5-year RFS of 64.8% (95% CI, 60.1% to 70.0%) for ColoPrint high-risk samples (n = 145).

Genes in the classifier were selected in an agnostic, data-driven way. Nevertheless, some of the selected genes have been shown to play a role in colon cancer biology (Appendix Table A2), coding for serine/threonine protein kinases, transcription factors, proteases, and membrane components. The gene ontology analysis (Babelomics software; http://babelomics.bioinfo.cipf.es/) revealed that the selected genes are involved in cell proliferation, transforming growth factor β pathway, immune response, and metabolism. One of the genes is *LAM3* (laminin-322), whose abnormal expression, in addition to its integrin receptors, is believed to promote invasion of colon, breast, and skin cancer cells. Moreover, *LAM3* and its protease degradation products may induce and/or promote tumor cell migration.[26] Another gene, *CTSC* (cathepsin C), has also been shown to be involved in invasion.

DISCUSSION

In this study, we present the development and validation of a gene expression signature that is associated with the risk of relapse in patients with stage II or III CRC. ColoPrint identifies two thirds of patients with stage II colon cancer who are at sufficiently low risk of recurrence who may be safely managed without adjuvant chemotherapy.

The unsupervised hierarchical clustering in three prognostic subtypes supports the underlying hypothesis that the transcripts of the primary tumors yield prognostic information. Of note, the molecular characteristics and percentage of patients in these three subtypes are

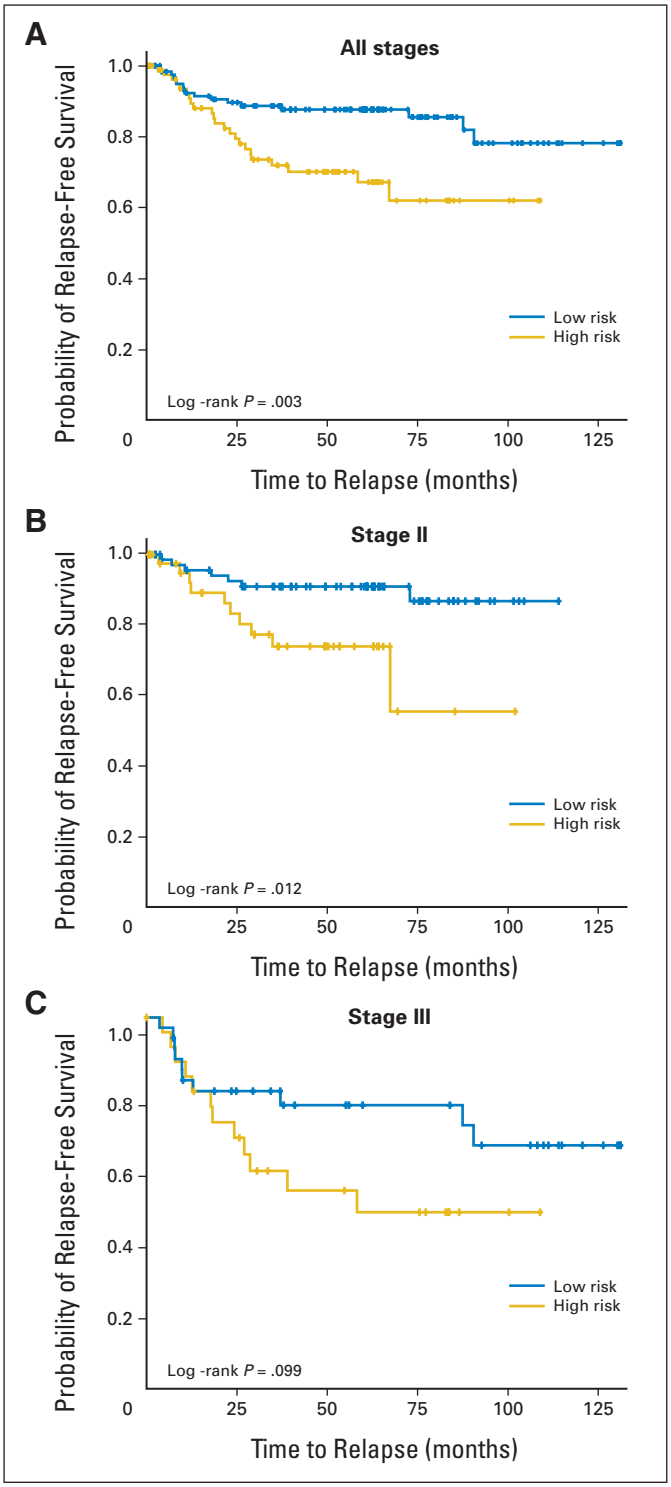

Fig 2. Kaplan-Meier analysis of relapse-free survival (RFS) in the validation set. (A) All stages, n = 206; 5-year RFS rate for low-risk patients was 87.6% (95% CI, 81.5% to 93.7%) and for high-risk patients was 67.2% (95% CI, 55.4% to 79.0%). (B) Stage II, n = 114; 5-year RFS rate for low-risk patients was 90.9% (95% CI, 84.0% to 97.8%) and for high-risk patients was 73.9% (95% CI, 59.2% to 88.6%). (C) Stage III, n = 62; 5-year RFS rate for low-risk patients was 78.2% (95% CI, 49.9% to 90.7%) and for high-risk patients was 47.2% (95% CI, 25.8% to 68.6%).

Table 3. Univariate Analysis for Relapse-Free Survival in Validation Set

Variable	*P*	HR	95% CI
All stages, N = 206			
ColoPrint, high *v* low risk	.005	2.51	1.33 to 4.73
Age, ≤ *v* > 70 years	.071	1.78	0.95 to 3.33
Localization, right *v* left	.576	0.82	0.43 to 0.16
Grade			
Baseline	.149	1	
Moderate *v* low		0.89	0.46 to 1.76
High *v* low		2	0.82 to 5.72
Sex, male *v* female	.739	1.12	0.58 to 2.14
No. of LNs assessed, continuous	.036	0.50	0.26 to 0.96
> 12 LNs assessed, binary	.036	0.50	0.26 to 0.96
Stage			
I *v* II	.004	0.21	0.03 to 1.59
Baseline = II		1	
III *v* II		2.36	1.25 to 4.47
pT			
Baseline = T2	.006	1	
T3 *v* T2		2.08	0.64 to 0.68
T4 *v* T2		6.74	1.74 to 26.11
pT, continuous	.003	2.8	1.41 to 5.54
pN			
Baseline	.000	1	
1-3 positive LNs *v* no positive LNs		1.88	0.88 to 4.01
> 3 positive LNs *v* no positive LNs		5.73	2.69 to 12.21
Chemotherapy, yes *v* no	.414	0.77	0.40 to 1.46
MSI-H, yes *v* no	.830	1.07	0.59 to 1.92
Lymphatic invasion, yes *v* no	.100	1.87	0.89 to 3.93
Venous invasion, yes *v* no	.101	2.20	0.86 to 5.62
Perineural invasion, yes *v* no	.651	1.58	0.22 to 11.54
Any invasion, yes *v* no	.051	1.93	1.00 to 3.76
Stage II only, n = 114			
ColoPrint, high *v* low risk	.017	3.34	1.24 to 9.00
Age, ≤ *v* > 70 years	.187	0.47	0.15 to 1.44
Localization, right *v* left	.73	0.82	0.15 to 2.46
Grade			
Baseline	.515		
Moderate *v* low		0.66	0.23 to 1.91
High *v* low		2.15	0.27 to 16.87
Sex, male *v* female	.175	2.17	0.71 to 6.67
No. of LNs assessed, continuous	.553	0.98	0.94 to 1.04
> 12 LNs assessed, binary	.776	0.86	0.30 to 2.44
pT, T4 *v* T3	.045	3.15	1.02 to 9.69
ASCO risk, high *v* low	.200	1.67	0.22 to 12.59
Chemotherapy, yes *v* no	.339	0.60	0.21 to 1.71
MSI-H, yes *v* no	.619	0.77	0.28 to 2.13
Lymphatic invasion, yes *v* no	.689	1.51	0.20 to 11.50
Venous invasion, yes *v* no	.496	2.02	0.27 to 15.41
Perineural invasion, yes *v* no	.237	3.41	0.45 to 26.03
Any invasion, yes *v* no	.209	2.23	0.64 to 7.80

Abbreviations: HR, hazard ratio; LN, lymph node; MSI-H, microsatellite instability-high; ASCO, American Society of Clinical Oncology.

reminiscent of the molecular CpG island methylation phenotype subtypes that are characterized by MSI, *BRAF* mutation, and methylation status.[27]

On the basis of gene expression information in the primary tumor, ColoPrint can assist in more accurately identifying the 25% to 35% of patients diagnosed with stage II disease who will experience a recurrence within 5 years after surgery. Our prognostic classifier identified 36.8% of the validation stage II subset as high-risk patients with

Table 4. Multivariate Analysis for Relapse-Free Survival in Validation Set

Variable	*P*	HR	95% CI
All stages, N = 206			
ColoPrint, high *v* low	.003	2.69	1.41 to 5.14
pT			
T2	.000		
T3 *v* T2	.038	0.19	0.04 to 0.91
T4 *v* T2	.960	1.05	0.19 to 5.88
Stage, continuous	.021	0.05	0.00 to 0.063
pN			
No positive LNs	.000		
1-3 positive LNs *v* no positive LNs	.327	1.52	0.66 to 3.52
> 3 positive LNs *v* no positive LNs	.000	5.97	2.62 to 13.63
No. of LNs assessed, continuous	.059		
Lymphatic, venous, or perineural invasion, any	.491		
Stage II only, n = 114			
ColoPrint, high *v* low	.018	3.29	1.24 to 8.83
pT, T4 *v* T3	.051	3.06	0.99 to 9.44

NOTE. Multivariate analysis includes only variables that were significant ($P < .05$) in the univariate analysis.
Abbreviations: HR, hazard ratio; LN, lymph node.

higher accuracy than the recommended clinical risk factors, irrespective of chemotherapy administration. Approximately two thirds of all patients analyzed received a low-risk classification, 91% of whom did not experience relapse. Chemotherapy was administered to 36% of these patients, where it was evenly distributed between ColoPrint high- and low-risk groups and had no influence in the global prognostic statistical analysis.

The suitability of gene expression profiles to identify high-risk patients with CRC has been proven in several independent studies.[15-21] Similar to what has been observed in the breast cancer field, these studies led to the construction of different gene signatures that may be secondary to differences in patient cohorts, technologic platforms, and data mining strategies.[12,28] Although signatures can often be validated in silico, the signature presented in this study is the first prognostic CRC profile, to our knowledge, that has been validated in independent patient series, using the same technology, gene set, and analytic approach.

The ColoPrint signature adds value to more conventional prognostic clinicopathologic factors. Routine standardization of genomic assessment, which includes tissue handling and processing, RNA extraction techniques, and the hybridization process, is a critical issue if it is to be used in the clinical setting.[29] Much progress has been made to establish high-quality standards for this new technology.[14] Of note, it is often overlooked that more conventional parameters, such as vascular invasion or a precise cutoff number of analyzed lymph nodes, may not be consistently recorded or evaluated in a significant proportion of tumors,[30] and a comparable standardization for clinical factors should be pursued.

Single molecular markers, such as loss of heterozygosity in 18q, MSI, thymidylate synthase expression, p53 or p21 expression, or *KRAS* or *BRAF* mutations, provide an additional means of characterizing individual tumors but are not routinely recommended for prognostic characterization.[31] MSI is the most extensively investigated and validated of these markers. A published meta-analysis showed that MSI is an independent prognostic predictor of improved survival and time to recurrence.[9] Data coming from the translational studies of the Quick and Simple and Reliable (QUASAR) trial and Pan-European Trials in Alimentary Tract Cancers (PETACC-3) have confirmed this observation in stage II disease.[10,11] However, this prognostic factor identifies only a small subgroup of low-risk patients.[23] In our validation data set, 8% of patients with known MSI status were MSI-H, and most of the patients (86%) were also identified as low risk by the ColoPrint classifier.

In this study, the prognostic classifier was validated in an independent patient set collected from a different country, with further international validation studies currently underway. Additionally, for use in routine clinical practice, the prognostic classifier was translated into a robust and standardized assay with stringent quality controls following guidelines of the National Committee on Clinical Laboratory Standards. Using the classifier in a clinical setting will provide more accurate information on the risk of recurrence compared with the use of conventional clinicopathologic criteria alone and can facilitate the selection of low-risk patients who can be spared chemotherapy.

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: Paul Roepman, Agendia (C); Iris Simon, Agendia (C); Christa Dreezen, Agendia (C); Annuska M. Glas, Agendia (C); Laura J. Van't Veer, Agendia (C) **Consultant or Advisory Role:** David Kerr, Agendia (C) **Stock Ownership:** Paul Roepman, Agendia; Iris Simon, Agendia; Annuska M. Glas, Agendia; Laura J. Van't Veer, Agendia **Honoraria:** None **Research Funding:** None **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Ramon Salazar, Paul Roepman, Gabriel Capella, Victor Moreno, Iris Simon, David Kerr, Laura J. Van't Veer, Rob Tollenaar

Financial support: Gabriel Capella, Victor Moreno, Laura J. Van't Veer, Rob Tollenaar

Administrative support: Iris Simon, Christa Dreezen, Adriana Lopez-Doriga, Sjoerd Bruin

Provision of study materials or patients: Ramon Salazar, Gabriel Capella, Victor Moreno, Cristina Santos, Corrie Marijnen, Johan Westerga, Sjoerd Bruin, Peter Kuppen, Cornelis van de Velde, Hans Morreau, Loes Van Velthuysen, Rob Tollenaar

Collection and assembly of data: Ramon Salazar, Paul Roepman, Gabriel Capella, Iris Simon, Adriana Lopez-Doriga, Cristina Santos, Corrie Marijnen, Johan Westerga, Sjoerd Bruin, Peter Kuppen, Cornelis van de Velde, Hans Morreau, Loes Van Velthuysen

Data analysis and interpretation: Ramon Salazar, Paul Roepman, Victor Moreno, Iris Simon, Christa Dreezen, Corrie Marijnen, Annuska M. Glas, Laura J. Van't Veer, Rob Tollenaar

Manuscript writing: Ramon Salazar, Iris Simon, Laura J. Van't Veer, Rob Tollenaar
Final approval of manuscript: Ramon Salazar, Paul Roepman, Gabriel Capella, Victor Moreno, Iris Simon, Christa Dreezen, Adriana Lopez-Doriga, Cristina Santos, Corrie Marijnen, Johan Westerga, Sjoerd Bruin, David Kerr, Peter Kuppen, Cornelis van de Velde, Hans Morreau, Loes Van Velthuysen, Annuska M. Glas, Laura J. Van't Veer, Rob Tollenaar

REFERENCES

1. O'Connell JB, Maggard MA, Ko CY: Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst 96:1420-1425, 2004

2. Quasar Collaborative Group, Gray R, Barnwell J, et al: Adjuvant chemotherapy versus observation in patients with colorectal cancer: A randomised study. Lancet 370:2020-2029, 2007

3. Sobrero A: Should adjuvant chemotherapy become standard treatment for patients with stage II colon cancer? For the proposal. Lancet Oncol 7:515-516, 2006

4. Köhne CH: Should adjuvant chemotherapy become standard treatment for patients with stage II colon cancer? Against the proposal. Lancet Oncol 7:516-517, 2006

5. 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

6. Van Cutsem EJ, Oliveira J, Kataja VV: ESMO Minimum Clinical Recommendations for diagnosis, treatment and follow-up of advanced colorectal cancer. Ann Oncol 16:i18-i19, 2005 (suppl 1)

7. 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

8. Le Voyer TE, Sigurdson ER, Hanlon AL, et al: Colon cancer survival is associated with increasing number of lymph nodes analyzed: A secondary survey of intergroup trial INT-0089. J Clin Oncol 21:2912-2919, 2003

9. Popat S, Hubner R, Houlston RS: Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 23:609-618, 2005

10. 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; abstr 4000)

11. Roth AD, Tejpar S, Yan P, et al: Stage-specific prognostic value of molecular markers in colon cancer: Results of the translational study on the PETACC 3-EORTC 40993-SAKK 60-00 trial. J Clin Oncol 27:169s, 2009 (suppl; abstr 4002)

12. Quackenbush J: Microarray analysis and tumor classification. N Engl J Med 354:2463-2472, 2006

13. van't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-536, 2002

14. Glas AM, Floore A, Delahaye LJ, et al: Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 7:278, 2006

15. Barrier A, Boelle PY, Roser F, et al: Stage II colon cancer prognosis prediction by tumor gene expression profiling. J Clin Oncol 24:4685-4691, 2006

16. 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

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

18. Garman KS, Acharya CR, Edelman E, et al: A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. Proc Natl Acad Sci U S A 105:19432-19437, 2008

19. Jiang Y, Casey G, Lavery IC, et al: Development of a clinically feasible molecular assay to predict recurrence of stage II colon cancer. J Mol Diagn 10:346-354, 2008

20. Lin YH, Friederichs J, Black MA, et al: Multiple gene expression classifiers from different array platforms predict poor prognosis of colorectal cancer. Clin Cancer Res 13:498-507, 2007

21. 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

22. Barrier A, Roser F, Boëlle PY, et al: Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. Oncogene 26:2642-2648, 2007

23. González-García I, Moreno V, Navarro M, et al: Standardized approach for microsatellite instability detection in colorectal carcinomas. J Natl Cancer Inst 92:544-549, 2000

24. Michiels S, Koscielny S, Hill C: Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 365:488-492, 2005

25. Kaiser S, Park YK, Franklin JL, et al: Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. Genome Biol 8:R131, 2007

26. Tsuruta D, Kobayashi H, Imanishi H, et al: Laminin-332-integrin interaction: A target for cancer therapy? Curr Med Chem 15:1968-1975, 2008

27. Poynter JN, Siegmund KD, Weisenberger DJ, et al: Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. Cancer Epidemiol Biomarkers Prev 17:3208-3215, 2008

28. Ein-Dor L, Kela I, Getz G, et al: Outcome signature genes in breast cancer: Is there a unique set? Bioinformatics 21:171-178, 2005

29. Cascinu S, Zaniboni A, Scartozzi M, et al: Molecular biology for stage II colorectal cancer: The jury is still out. J Clin Oncol 25:2861; author reply 2862-2863, 2007

30. Morris EJ, Maughan NJ, Forman D, et al: Who to treat with adjuvant therapy in Dukes B/stage II colorectal cancer? The need for high quality pathology. Gut 56:1419-1425, 2007

31. Locker GY, Hamilton S, Harris J, et al: ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol 24:5313-5327, 2006

