







Contribution of Immunoscore and Molecular Features to Survival Prediction in Stage III Colon Cancer

Frank A. Sinicrope ¹, Qian Shi², Fabienne Hermitte ³, Tyler J. Zemla², Bernhard Mlecnik ^{4,5}, Al B. Benson⁶, Sharlene Gill⁷, Richard M. Goldberg ⁸, Morton S. Kahlenberg⁹, Suresh G. Nair¹⁰, Anthony F. Shields ¹¹, Thomas C. Smyrk¹, Jerome Galon⁴, Steven R. Alberts ¹

¹Division of Oncology, Department of Medicine, Mayo Clinic, Rochester, MN, USA; ²Alliance Statistics and Data Center, Rochester, MN, USA; ³HaliDx, Marseille, France; ⁴INSERM, UMRS 1138, Laboratory of Integrative Cancer Immunology, Université Paris Descartes, Paris, France; ⁵Inovarian, Paris, France; ⁶Northwestern University, Chicago, IL, USA; ⁷British Columbia Cancer Agency- Vancouver Cancer Centre, Vancouver, BC, Canada; ⁸West Virginia Cancer Institute, Morgantown, WV, USA; ⁹Surgical Oncology Associates of South Texas, San Antonio, TX, USA; ¹⁰Lehigh Valley Hospital, Allentown, PA, USA and ¹¹Karmanos Cancer Institute, Detroit, MI, USA

*Correspondence to: Frank A. Sinicrope, MD, Division of Oncology, Department of Medicine, Mayo Clinic, 200 1st St SW, Rochester, MN 55902, USA (e-mail: sinicrope.frank@mayo.edu).

Abstract

Background: The American Joint Committee on Cancer staging and other prognostic tools fail to account for stage-independent variability in outcome. We developed a prognostic classifier adding Immunoscore to clinicopathological and molecular features in patients with stage III colon cancer. **Methods:** Patient ($n = 559$) data from the FOLFOX arm of adjuvant trial NCCTG N0147 were used to construct Cox models for predicting disease-free survival (DFS). Variables included age, sex, T stage, positive lymph nodes (+LNs), N stage, performance status, histologic grade, sidedness, KRAS/BRAF, mismatch repair, and Immunoscore (CD3⁺, CD8⁺ T-cell densities). After determining optimal functional form (continuous or categorical) and within Cox models, backward selection was performed to analyze all variables as candidate predictors. All statistical tests were two-sided. **Results:** Poorer DFS was found for tumors that were T4 vs T3 (hazard ratio [HR] = 1.76, 95% confidence interval [CI] = 1.19 to 2.60; $P = .004$), right- vs left-sided (HR = 1.52, 95% CI = 1.14 to 2.04; $P = .005$), BRAF^{V600E} (HR = 1.74, 95% CI = 1.26 to 2.40; $P < .001$), mutant KRAS (HR = 1.66, 95% CI = 1.08 to 2.55; $P = .02$), and low vs high Immunoscore (HR = 1.69, 95% CI = 1.22 to 2.33; $P = .001$) (all $P < .02$). Increasing numbers of +LNs and lower continuous Immunoscore were associated with poorer DFS that achieved significance (both $P < .0001$). After number of +LNs, T stage, and BRAF/KRAS, Immunoscore was the most informative predictor of DFS shown multivariately. Among T₁₋₃ N₁ tumors, Immunoscore was the only variable associated with DFS that achieved statistical significance. A nomogram was generated to determine the likelihood of being recurrence-free at 3 years. **Conclusions:** The Immunoscore can enhance the accuracy of survival prediction among patients with stage III colon cancer.

Colorectal cancer (CRC) is the second leading cause of cancer-related death in the United States (1). The American Joint Committee on Cancer (AJCC) staging system is the most widely used and includes limited disease characteristics (ie, tumor stage, nodal status, and metastases [TNM]) to develop stage-specific prognoses (2). Within existing staging systems, however, there remains considerable intrastage variability in clinical outcome (3). AJCC staging and web-based calculators (4) do not include intratumoral features of immune infiltration or molecular characteristics and DNA mismatch repair status that are associated with prognosis in colon cancer and that may help to refine prognostication. Furthermore, current staging systems and prognostic tools were developed using observational data

with the potential for biases inherent to nonrandomized studies. In the era of personalized medicine, refinement of prognostic accuracy may lead to better treatment planning and patient-physician communication.

The host antitumor immune response includes cytotoxic and memory T cells, and scores combining CD3⁺ and cytotoxic CD8⁺ T-cell densities in core tumor (CT) and at the invasive margin (IM), known as Immunoscore (5,6), were associated with colon cancer survival and may enable risk stratification beyond AJCC TNM staging (6–10). However, prior studies of the Immunoscore consist primarily of retrospective case series of pooled CRC stages with varying treatment and follow-up. Prognosis may vary by tumor stage given differences

Received: October 21, 2019; Revised: January 17, 2020; Accepted: March 24, 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

in immune composition and immunobiology (11), as well as differences in the impact of molecular alterations as shown for DNA mismatch repair (MMR) status (12). Accordingly, evaluation of the Immunoscore in a single tumor stage and uniform treatment with a standard chemotherapy regimen (13) is an important and unmet need.

To address a major limitation of prior research, we sought to identify a stage III-specific risk classification for prognosis. Accordingly, we constructed prediction models for disease-free survival (DFS) that incorporate Immunoscore, BRAF^{V600E}, KRAS, and MMR status in addition to clinicopathological features that are relevant to patient outcomes after adjuvant treatment. Models were constructed for stage III patients treated with FOLFOX in a phase 3 randomized clinical trial, and a nomogram was generated that enables patient-specific predictions of DFS.

Methods

Study Population

Patients with curatively resected stage III colon adenocarcinomas were participants in a phase 3 randomized trial of adjuvant FOLFOX ± cetuximab (NCCTG N0147; Clinicaltrials.gov Identifier: NCT00079274) (14). Clinical and pathological characteristics are shown in Table 1 (see "Results"). Study patients were randomly selected from the FOLFOX-alone arm and were representative of all patients on this arm. Data were available on 559 cases with Immunoscore values that met prespecified quality control criteria. Tumor recurrence occurred in 164 (29.3%) patients; 137 died during follow-up (median = 77 months). Median number of examined lymph nodes (LNs) was 18. Primary tumor site was categorized as right-sided if proximal to splenic flexure. The study was approved by the Mayo Clinic Institutional Review Board and by the Alliance for Clinical Trials in Oncology. Each participant signed an Institutional Review Board-approved, protocol-specific informed consent.

Immune Marker Analysis

Immunostaining for CD3⁺ and CD8⁺ T cells and for CD20⁺ B lymphocytes was performed on archival sections cut from one tumor block containing tumor and IM from each patient and selected by one gastrointestinal pathologist (TCS). Sections were incubated at 37°C with primary antibodies: rabbit monoclonal antihuman CD3 (MRQ-39, Cell Marque, Rocklin, CA), mouse antihuman monoclonal antibodies against CD8⁺ (C8/144B, Cell Marque), and CD20⁺ (MO755, clone L26, DAKO). Counterstained slides were scanned, and digital images of stained tissue sections were obtained at 10x magnification and 0.45 μm/pixel resolution (NanoZoomer XR, Hamamatsu, Japan). Image analysis software (Immunoscore Analyzer, HalioDx, Marseille, France) was used for automatic tissue detection (CT, healthy nonepithelial tissue, and epithelium) and to quantify the density of stained immune cells by number of cells per mm² in both CT and IM (5,6,15). IM was defined as a region of 360 μm width on each side of the border between malignant cells and peritumoral stroma. CD3⁺ and CD8⁺ T-cell densities were converted into Immunoscore with predefined cutoffs (5). Immunoscore uses standardized percentile values (0%–100%), and the algorithm categorizes the continuous Immunoscore into five groups

(0, 1, 2, 3, and 4). A predefined three-level classification groups Immunoscore into low (0–1), intermediate (2), and high (3–4) groups, whereas the two-level classification uses predefined cutoffs corresponding to low (0–1, mean percentile 0%–25%) and high Immunoscore (2–4; mean percentile >25%–100%). Results were analyzed blinded to clinical outcome data. Samples were excluded from the analysis if counts were missing from a tumor region or if staining intensity was deemed low.

Tumor DNA MMR status was determined by immunohistochemical analysis of MMR proteins (MLH1, MSH2, MSH6) (16); loss of at least one MMR protein indicated deficient MMR (dMMR). BRAF c.1799T>A (V600E) mutation (exon 15) was detected using a multiplex allele-specific polymerase chain reaction-based assay. KRAS (exon 2) was analyzed using the DxS Mutation Test Kit KR-03/04 (DxS) assessing for seven mutations in codons 12 and 13. All authors had access to study data and reviewed and approved the final manuscript.

Statistical Analysis

Of the patients, 527 with all variables of interest were included in construction of clinical prediction models for DFS (defined as time to recurrence or death, whichever occurred first). The optimal functional form (ie, continuous or categorical) was determined for each variable: age (continuous), sex (male vs female), performance status (PS; 0 vs 1), tumor grade (high vs low), T stage (T1/T2 vs T3 vs T4), number of positive LNs, and tumor site (left vs right). Additional variables included CD3⁺ CT, CD3⁺ IM, CD8⁺ CT, CD8⁺ IM (all continuous), Immunoscore (continuous), MMR (dMMR vs proficient MMR), and BRAF/KRAS (WT/WT vs MUT/WT vs WT/MUT). The proportional hazards assumption was verified using methods of Grambsch and Therneau (17). For each categorical variable, a Kaplan-Meier curve was created comparing all levels and with determination of whether any levels could be combined/excluded.

Continuous variables were modeled using restricted cubic splines to assess possible nonlinearity of their effects and to select appropriate degrees of freedom for modeling (18). For each continuous variable, it was determined whether a spline term was necessary in the model vs a linear term. If a nonlinear term was necessary, additional knots were added as necessitated by visual inspection of spline plots and formal testing for statistical significance (beyond three knots default at initial stage) (18). A multivariable Cox regression model was constructed with all variables, each in optimal functional form. A backward elimination procedure (*P* value of .05 as threshold for staying in model) was applied at each iteration to identify the most important independent covariates to predict DFS. Pairwise interactions were tested for covariates remaining in the final Cox model and included in final models if statistical significance (*P* < .05 per Wald test) was achieved, although none were statistically significant. Final models included all relevant main effects and pairwise interactions. Fitted models were used to construct a nomogram for 3-year DFS probability (19,20). All statistical tests were two-sided. Analyses and figures were produced using package rms [R statistical software version 2.15.2 (19)]. A *P* value less than .05 was considered statistically significant.

Study data collection and statistical analyses were conducted by the Alliance Statistics and Data Center.

Table 1. Clinicopathological and molecular features of stage III colon carcinomas treated with adjuvant FOLFOX in a phase III trial

Variables	Total (n = 559)
Age	
Median	59.0
Range	(21.0–83.0)
Sex, No. (%)	
Female	273 (48.8)
Male	286 (51.2)
T stage, No. (%)	
T1 or T2	82 (14.7)
T3	416 (74.4)
T4	61 (10.9)
N stage, No. (%)	
1–3 (N1)	328 (58.7)
≥4 (N2)	231 (41.3)
No. of positive LNs	
Median	3.0
Range	(1.0–51.0)
Histologic grade, No. (%)	
High	148 (26.5)
Low	411 (73.5)
Performance score, No. (%)	
0	438 (78.4)
1	118 (21.1)
2	3 (0.5)
Tumor location, No. (%)*	
Right	255 (46.1)
Left	298 (53.9)
BRAF/KRAS, No. (%)*	
WT/WT	292 (54.0)
WT/MUT	175 (32.3)
MUT/WT	74 (13.7)
MMR, No. (%)*	
pMMR	499 (90.1)
dMMR	55 (9.9)

*Missing data. dMMR = deficient MMR; LN = lymph node; MUT = mutant; pMMR = proficient mismatch repair; WT = wild-type.

Results

Patient Characteristics

Characteristics of the study population are shown in [Table 1](#). Among stage III tumors analyzed, 10.9% were T4, 41.3% were N2, 13.7% had mutated BRAF(V600E), 32.3% had mutated KRAS, and 9.9% showed dMMR. Representative hematoxylin-stained tissue sections showing CD3⁺, CD8⁺ T cells, and CD20⁺ B cells are provided ([Figure 1](#)). CT and IM are shown with demarcation of the IM separating malignant glands/cells from peritumoral stroma.

Validation of the Consensus Immunoreactivity for DFS

We validated the two-level categorical Immunoreactivity ([Figure 2](#)) whose prognostic impact was previously shown in an international validation study in TNM stage I–III colon cancers ([5](#)). For consistency with prior work, the associations with DFS are also shown for Immunoreactivity categorical five- and three-level variables ([Supplementary Figure 1](#), A and B, available online). When tumors were categorized into predetermined low (0–1) and high (2–4) groups ([21](#)), a low Immunoreactivity was associated with a

statistically significant and poorer DFS (hazard ratio [HR]_{adj} = 1.69, 95% confidence interval [CI] = 1.22 to 2.33; *P*_{adj} = .001) after adjusting for age, tumor location, T and N stage, BRAF/KRAS, and MMR status (3-year DFS for low vs high: 66.6% vs 82.6%) ([Figure 2](#)).

Association of Clinical and Molecular Variables with DFS

We determined the optimal functional form for all variables included in the analysis and their relationship to DFS univariately ([Table 2](#); [Supplementary Figures 2 and 3](#), available online). Sex, histologic grade, Eastern Cooperative Oncology Group PS, tumor site, and MMR were evaluated as two-level categorical variables. Of these variables, only primary tumor site was statistically significant univariately ([Table 2](#)). Patients with right-sided tumors had a statistically significant and worse DFS compared with left-sided tumors (HR = 1.52, 95% CI = 1.14 to 1.14; *P* = .005). Patients with T4 tumors had the poorest DFS relative to T3 (HR = 1.76, 95% CI = 1.19 to 2.60; *P* = .004) or T1/2 (*P* < .001) tumors that achieved statistical significance ([Table 2](#)). T3 tumors also had worse DFS vs T1/T2 (HR = 2.95, 95% CI = 1.60 to 5.44; *P* < .001). Accordingly, T stage was analyzed as a three-level variable in subsequent model selection procedures. In univariate analysis, Immunoreactivity was analyzed as a continuous variable and higher scores (10-percentile increased) were associated with a statistically significant and better DFS (HR = 0.87, 95% CI = 0.81 to 0.93; *P* < .001) ([Table 2](#)). MMR was not prognostic (*P* = .62), whereas tumors harboring BRAF^{V600E} (HR = 1.74, 95% CI = 1.26 to 2.40; *P* < .001) or mutant KRAS (HR = 1.66, 95% CI = 1.08 to 2.55; *P* = .02) were associated with a statistically significant and shorter DFS compared with tumors lacking either mutation ([Table 2](#); [Supplementary Figure 2](#), E and F, available online).

To enhance model precision, the effect of each continuous variable was tested for a potential nonlinear effect with restricted cubic splines within Cox models. There was no statistical evidence of a nonlinear effect for age, CD3⁺ CT, CD8⁺ CT, or Immunoreactivity (ie, nonlinear term *P* > .05) ([Supplementary Figure 3](#), available online). However, there were statistically significant nonlinear effects for positive LNs (nonlinear term *P* < .001, three knots), CD8⁺ IM (nonlinear term *P* = .001, four knots), and CD3⁺ IM (nonlinear term *P* < .001, four knots). Although the association of age with DFS was not statistically significant (*P* = .92), an increasing number of positive LNs were associated with shorter DFS that achieved statistical significance (*P* < .001) ([Supplementary Figure 3B](#), available online). CD3⁺ and CD8⁺ densities, both in CT and IM, were each inversely associated with patient DFS that was statistically significant ([Supplementary Table 1](#) and [Supplementary Figure 3](#), C–F, available online). Similarly, Immunoreactivity (2.5% steps) was inversely associated with DFS and achieved statistical significance (*P* < .001) ([Supplementary Table 1](#) and [Supplementary Figure 3G](#), available online). In contrast to T-cell markers, CD20⁺ B-cell density did not display a statistically significant association with DFS (data not shown) and was therefore not included in multivariable models.

Identification of Candidate Predictors of Survival in Multivariable Models

A multivariable final model for DFS is presented in [Table 3](#). All variables in their optimal functional form were included as candidate predictors in the initial full model. A backward selection

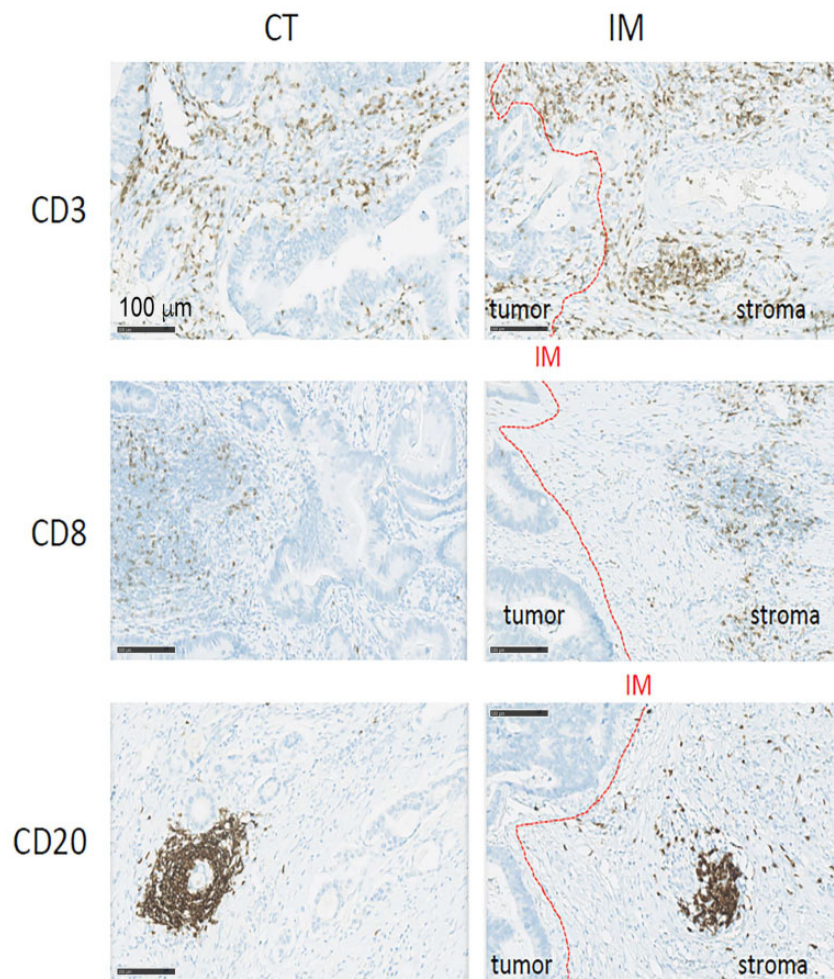


Figure 1. Representative hematoxylin-stained tissue sections showing CD3⁺ and CD8⁺ T cells and CD20⁺ B cells in stage III human colon cancers. The core tumor (CT) and the invasive margin (IM) are shown with the demarcation of the IM (red line; see "Methods") separating malignant glands/cells from peritumoral stroma.

was then performed (P value of .05 as threshold for remaining in the model at each iteration) to arrive at the final model that included variables that showed the strongest independent association with DFS (ie, number of positive LNs, T stage, BRAF/KRAS, and Immunoscore). Based on adjusted hazard ratios (Table 3) and relative risk contributions (Figure 3), each variable in the final model for DFS was deemed clinically significant. Analysis of the relative importance of all variables in the initial multivariable model revealed that the number of positive LNs (43.1%) had the largest impact on DFS risk followed by T stage (18%), BRAF/KRAS status (16.1%), and then Immunoscore (14.9%) (Figure 3). As an internal validation, we performed a bootstrap analysis of our prediction model. We found the bootstrap hazard ratios to be very similar to those shown in the final multivariate model (Table 3) with P values approaching statistical significance, as expected, in a Cox model. For example, the bootstrap analysis yielded hazard ratios for the continuous Immunoscore of 0.9 (95% CI = 0.83 to 0.97), 1.85 (95% CI = 1.30 to 2.64) for mutant KRAS, and 2.45 (95% CI = 1.22 to 4.93) for T stage (T3 vs T1/2). These data serve to internally validate our findings for the Immunoscore in our study cohort.

The final model was then used to generate a nomogram to estimate DFS rates for individual patients in clinical practice. The

nomogram, inclusive of the Immunoscore (2.5% steps), assigns points to each variable (depending on their results) to arrive at an estimated 3-year DFS rate (Figure 4). Using this nomogram, some patients believed to have a low risk of relapse are found to be at high risk based on inclusion of Immunoscore and/or mutated BRAF/KRAS. Conversely, some patients believed to be at high risk for relapse can be reclassified as having a lower risk of relapse. As an example, a patient with four positive LNs, T3, mutant BRAF, and an Immunoscore of 30% would have 226 total points corresponding to a 3-year DFS of 64%, as determined using the nomogram (Figure 4). If this same patient had an Immunoscore of 70%, then the DFS rate would be 75% (193 total points).

Categorization of stage III patients into low-risk ($T_{1-3}N_1$) and high-risk (T_4 or N_2) groups is routinely used to guide the recommended duration of adjuvant FOLFOX or CAPOX treatment in clinical practice (22). Using the same modeling procedure as done for the main analysis on the low-risk group ($n=296$) revealed that only Immunoscore remained statistically significant (HR = 0.87, 95% CI = 0.78, 0.98; $P = .02$ per (10-percentile increase; data not shown). Thus, the Immunoscore is the variable that is most strongly predictive of DFS among patients in the low-risk group. Similarly modeling of the high-risk group ($N=249$) revealed that Immunoscore [HR 0.89, 95% CI 0.81, 0.97,

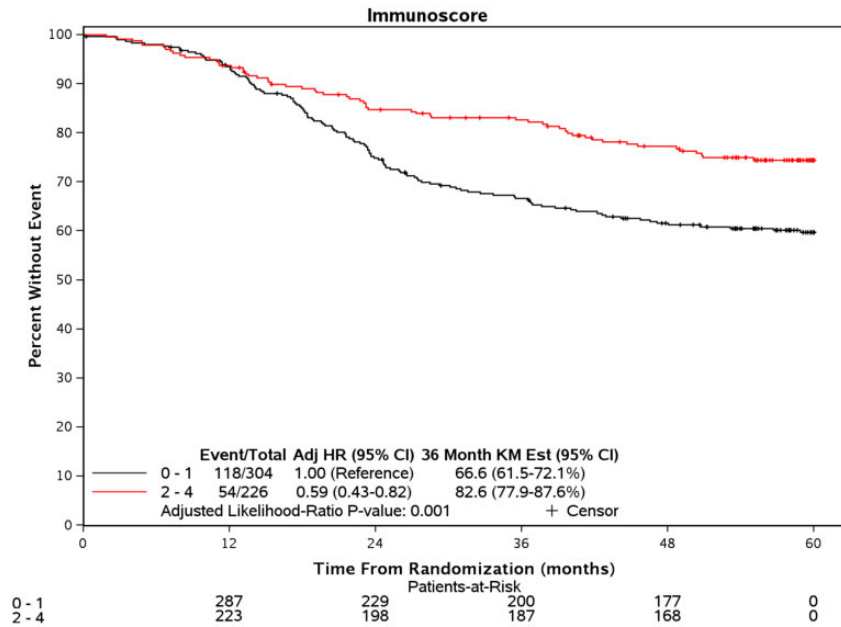


Figure 2. Association of the two-level categorical Immunoscore with patient 3-year disease-free survival (DFS) in 559 patients with stage III colon cancers as shown in a Kaplan-Meier (KM) plot. Immunoscore was dichotomized into predetermined low (0-1) vs high (2-4) values, as evaluated in an international validation study (5). Adj HR = adjusted hazard ratio; CI = confidence interval; Est = estimated.

Table 2. Univariate associations of clinical, immune, and molecular features (in best functional form) with DFS

Variable	No.	Type	Best functional form	HR (95% CI)	P*
Age	559	Continuous	Linear (5-year increase)	1.00 (0.94 to 1.07)	.92
Number of positive LN	559	Continuous	Spline (3 knots)	Not applicable	<.001
CD3 ⁺ CT	558	Continuous	Linear (10-percentile increase)	0.86 (0.79 to 0.95)	.002
CD3 ⁺ IM	542	Continuous	Spline (4 knots)	Not applicable	<.001
CD8 ⁺ CT	558	Continuous	Linear (10-percentile increase)	0.78 (0.63 to 0.96)	.17
CD8 ⁺ IM	542	Continuous	Spline (4 knots)	Not applicable	<.001
Immunoscore (2.5% steps)	559	Continuous	Linear (10-percentile increase)	0.87 (0.81 to 0.93)	<.001
Sex	559	Categorical	2 levels: male vs female	0.97 (0.73 to 1.30)	.86
T stage	559	Categorical	3 levels:		
			T4 vs T1/T2	5.19 (2.61 to 10.32)	<.001
			T3 vs T1/T2	2.95 (1.60 to 5.44)	<.001
			T4 vs T3	1.76 (1.19 to 2.60)	.004
Histology	559	Categorical	2 levels: high vs low	1.24 (0.90 to 1.71)	.19
Tumor site	553	Categorical	2 levels: left vs right	0.66 (0.49 to 0.88)	.005
BRAF/KRAS	541	Categorical	3 levels:		
			BRAF MUT vs double WT	1.74 (1.26 to 2.40)	<.001
			KRAS MUT vs double WT	1.66 (1.08 to 2.55)	.02
			KRAS MUT vs. BRAF MUT	0.96 (0.62 to 1.47)	.84
MMR status	554	Categorical	2 levels: dMMR vs pMMR	0.88 (0.53 to 1.47)	.62
Performance score	556	Categorical	2 levels: 1 vs 0 (PS = 2 [n = 3] excluded)	1.17 (0.83 to 1.66)	.37

*Two-sided Wald χ^2 P value. CI = confidence interval; CT = core tumor; DFS = disease-free survival; dMMR = deficient mismatch repair; HR = hazard ratio; IM = invasive margin; LN = lymph node; MMR = mismatch repair; MUT = mutant; pMMR = proficient mismatch repair; PS = performance status; WT = wild-type.

$p=0.01$ (per 10-percentile increase) along with BRAF^{v600E} [HR 1.81, 95%, CI 1.10, 2.97, $p=0.02$], and mutant KRAS [HR 1.76, 95%, CI 1.15, 2.71, $p=0.01$] remained statistically significant.

Discussion

Given considerable within-stage heterogeneity in patient prognosis, we developed a prognostic classifier in patients with stage III colon cancers treated with adjuvant FOLFOX. We

constructed multivariate Cox models that condense multiple risk variables that included established clinical, molecular, and immune features into a final model. Although T and N stages are established prognostic variables (13), the relative contributions of other variables such as Immunoscore and BRAF/KRAS mutational status to patient prognosis are relatively unknown. BRAF and KRAS were analyzed as a combined variable because mutations in these genes are nearly always mutually exclusive within a given tumor (23). We found quite similar relative

Table 3. Final multivariable Cox model for DFS

Characteristics	Adjusted HR (95% CI)	P*
Immunoscore, 10-percentile increase	0.90 (0.87 to 0.94)	.004
No. of positive LN	Not applicable	<.001
T stage		.002
T1 or T2	1.00 (Referent)	
T3	2.40 (2.14 to 2.68)	.009
T4	3.62 (2.50 to 5.26)	<.001
BRAF/KRAS		.001
WT/WT	1.00 (Referent)	
WT/MUT	1.84 (1.56 to 2.18)	<.001
MUT/WT	1.56 (1.24 to 1.95)	.05

*Two-sided Wald χ^2 P value. LN = lymph node; CI = confidence interval; DFS = disease-free survival; HR = hazard ratio; MUT = mutant; WT = wild-type.

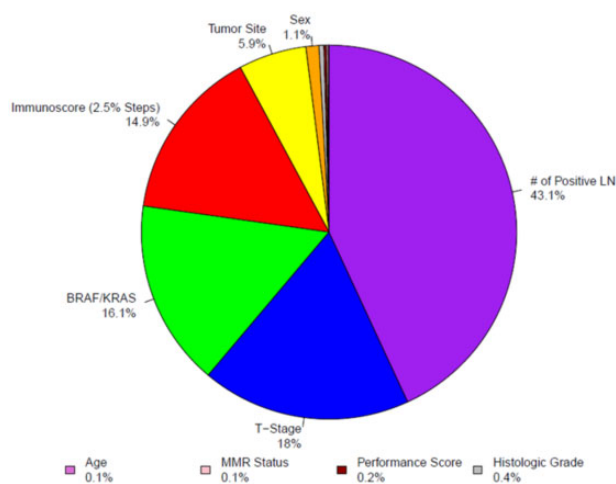


Figure 3. Relative contributions (%) of the 10 variables included in the multivariable Cox model to disease-free survival in patients with stage III colon cancer treated with adjuvant FOLFOX chemotherapy. LN = lymph node.

contributions of T stage (18%), BRAF/KRAS (16.1%), and Immunoscore (14.9%) to patient DFS. Although Immunoscore has been shown to provide independent prognostic information beyond clinicopathologic variables (6,7), we emphasize our novel findings in a single tumor stage (stage III) with molecular analysis and uniform treatment compared with prior studies that have included pooled tumor stages and variable treatment.

Decision-making for the duration of adjuvant chemotherapy in stage III disease uses T- and N-based high- and low-risk categories (13,22). Among IDEA trial-defined low-risk (T_{1-3} N_1) patients (22), Immunoscore was the strongest predictor of DFS because it remained the only variable in the final multivariate model after backward selection. In the overall population, we derived a nomogram from fitted model data whereby the user can input patient characteristics to obtain predicted probabilities of being recurrence-free and alive at 3 years beyond treatment initiation with confidence intervals. Three-year DFS is highly predictive of patient overall survival and is used as the study endpoint in adjuvant chemotherapy trials in colon cancers (19,20). The nomogram enables a more accurate and personalized DFS rate than with T, N alone, and accordingly, it can enhance prognostication and improve physician-patient communication with implications for treatment of individual patients.

Immunoscore reflects the interaction between the tumor microenvironment and the host immune system (24), and specific chemokines and adhesion molecules play important roles in determining the density of intratumoral T-cell densities (25). Analysis of Immunoscore as a continuous variable revealed that patients who had tumors with higher Immunoscores had better DFS that achieved statistical significance, whereas those with lower Immunoscore were more likely to suffer disease recurrence or death. Furthermore, analysis of the categorical Immunoscore using a prespecified cutoff was shown to validate the two-level consensus Immunoscore in our clinical trial cohort that was recently shown to predict clinical outcome in TNM stage I-III colon cancers in an international validation study (5). In this international study (5), a multivariable stratified Cox model including MMR/microsatellite instability status and the Immunoscore found that MMR was not a statistically significant factor for DFS (or overall survival) and was dependent on Immunoscore, as also shown in our study. Accordingly, the beneficial effect of dMMR/microsatellite instability high status was interpreted as mainly related to its capacity to induce strong antitumor immunity (ie, a high Immunoscore) (5). We observed differential outcome of the Immunoscore by primary tumor sidedness that achieved statistical significance as a prognostic factor in univariate analyses; however, it was not a component of the final multivariable risk model. This finding suggests that the prognostic value of tumor location is captured by measurement of the Immunoscore. Differences in the biology of colon cancer originating in the left vs right colon have been shown by multiomics including differentially expressed genes, miRNAs, and methylation changes (26), and the interplay of tumor sidedness with tumor immunity warrants further examination. Interestingly, intratumoral immune densities may be influenced by the gut microbiome in CRC, as high levels of *Fusobacterium nucleatum* DNA increased from rectum to cecum, and *F. nucleatum* quantity was inversely associated with CD3⁺ T-cell density (27, 28).

Strengths of our study include the clinical trial cohort of patients of uniform stage and treatment with meticulous collection of outcome data. Advantages inherent to randomized trials include balance of patient characteristics across arms, controlled sources of bias, and rigorous patient follow-up that leads to improved prediction modeling. Although our analysis was limited to KRAS, mutations in NRAS and HRAS occur in fewer than 5% of CRCs (29). A potential limitation is the generalizability of results to patients who might not resemble those eligible for enrollment in the clinical trial. Our final model warrants external validation in an independent cohort of FOLFOX-treated, stage III patients. Because all patients received adjuvant chemotherapy, we were unable to examine the predictive impact of covariates for chemotherapy response. Relevant to this issue are data indicating that oxaliplatin may increase cytotoxic T-cell infiltration and may induce immunogenic cell death (30).

In conclusion, our data go beyond anatomic tumor staging to demonstrate that after BRAF/KRAS status, the Immunoscore is the most informative contributor to the prediction of patient DFS. Our data also serve to validate the categorical Immunoscore to prognostically stratify stage III patients treated with adjuvant FOLFOX. We developed a nomogram that can enhance patient prognostication and physician-patient communication. Furthermore, the nomogram has the potential to replace the current list of stratification factors used to balance baseline risk across treatment arms at clinical trial randomization. We propose that model construction and generation of the nomogram be extended to stage II disease, where enhanced risk

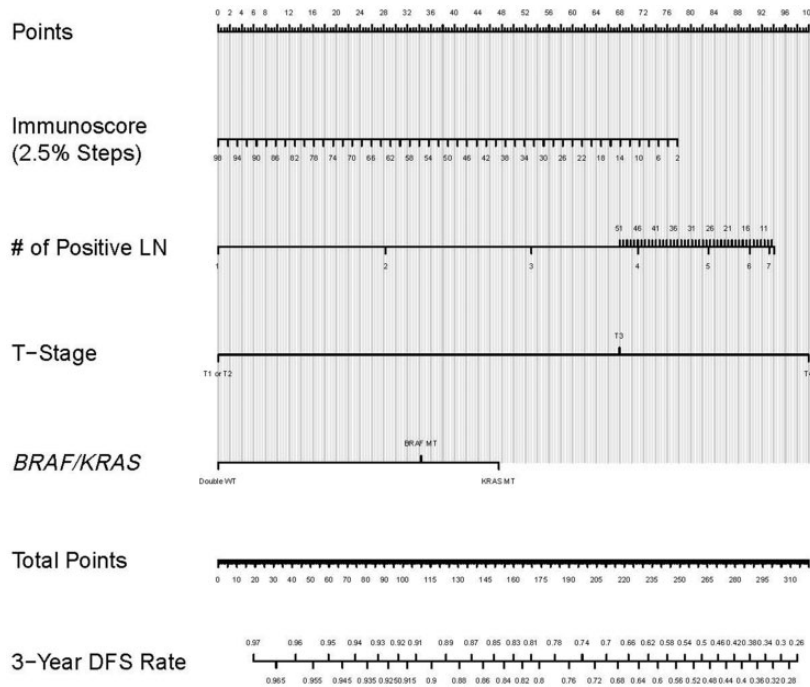


Figure 4. Nomogram for 3-year disease-free survival (DFS). An example of an individual patient's prediction may be obtained from the nomogram as follows. First, risk points associated with each variable are obtained via vertical translation of the patient's variable value e.g., Immunoscore, to the scale labeled "Points" in the nomogram (i.e., having Immunoscore of 46 contributes 42 points to recurrence/death risk). Next, the points associated with each variable value for the patient are totaled across the variables. This total is then located on the scale "Total Points" (e.g., 185) and then vertically mapped to obtain the 3-year DFS probability. As an example, a patient with an Immunoscore of 46.0%, one positive LN, T3, and mutant KRAS would have 185 total points corresponding to a 3 yr DFS of 78.0%.

stratification is needed to guide the selection of patients for adjuvant chemotherapy.

Funding

This work was supported by the NCCTG N0147 trial, and related correlative science was supported, in part, by grants from the US National Cancer Institute (NCI) under award numbers R01 CA210509-01A1 (to FAS), U10CA025224 (to North Central Treatment Group), and U24CA114740 (NCCTG Biospecimen Resource). Intergroup participation was made possible by NCI awards U10CA180888, UG1CA233163, and U10CA180835. The study was also supported in part by funds from Bristol-Myers Squibb, ImClone Systems Inc, Sanofi Aventis, and Pfizer. Bristol-Myers Squibb provided cetuximab to the NCCTG for conduct of the study. French National Institute of Health and Medical Research (INSERM), Eli Lilly and Company®, Pfizer®, and Bristol-Myers Squibb®.

Notes

Role of the funders: The funders had no role in the design of the study or in the data collection, analysis, or interpretation. Furthermore, the funders did not participate in the writing of the manuscript or in the decision to submit the manuscript for publication.

Conflicts of interest: JG and FH are co-founders of HaliDx. The other authors report no relevant conflicts of interest.

Acknowledgments: The authors wish to recognize the substantial contributions of the late Dan Sargent, PhD, to correlative studies, including immune biomarkers, in the NCCTG N0147 trial. We acknowledge Aurelie Cateau and Graham Poage for their careful reading of the manuscript. Appreciation is extended to Diane Xamath-Dos for her secretarial assistance.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; 69(1):7–34.
2. Weiser MR. AJCC 8th edition: colorectal cancer. *Ann Surg Oncol.* 2018;25(6): 1454–1455.
3. Nagtegaal ID, Quirke P, Schmol HJ. Has the new TNM classification for colorectal cancer improved care? *Nat Rev Clin Oncol.* 2012;9(2):119–123.
4. Renfro LA, Grothey A, Xue Y, et al. ACCENT-based web calculators to predict recurrence and overall survival in stage III colon cancer. *J Natl Cancer Inst.* 2014;106(12):dju333.
5. Pages F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet.* 2018;391(10135):2128–2139.
6. Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* 2006;313(5795):1960–1964.
7. 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(25):2654–2666.
8. Naito Y, Saito K, Shiiba K, et al. CD8+ T cells infiltrated within cancer cell nests as a prognostic factor in human colorectal cancer. *Cancer Res.* 1998; 58(16):3491–3494.
9. Prall F, Duhrop K, Weirich V, et al. Prognostic role of CD8+ tumor-infiltrating lymphocytes in stage III colorectal cancer with and without microsatellite instability. *Hum Pathol.* 2004;35(7):808–816.
10. Rozek LS, Schmit SL, Greenon JK, et al. Tumor-infiltrating lymphocytes, Crohn's-like lymphoid reaction, and survival from colorectal cancer. *J Natl Cancer Inst.* 2016;108(8):djw027.
11. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity.* 2013;39(4):782–795.

12. Price TJ, Karapetis CS, Joanne Y, et al. Outcomes for metastatic colorectal cancer (mCRC) based on microsatellite instability. *J Clin Oncol*. 2018;36(4)(suppl):759–759.
13. Benson AB, Venook AP, Al-Hawary MM, et al. NCCN guidelines insights: colon cancer, version 2.2018. *J Natl Compr Canc Netw*. 2018;16(4):359–369.
14. Alberts SR, Sargent DJ, Nair S, et al. Effect of oxaliplatin, fluorouracil, and leucovorin with or without cetuximab on survival among patients with resected stage III colon cancer: a randomized trial. *JAMA*. 2012;307(13):1383–1393.
15. Angell HK, Gray N, Womack C, Pritchard DI, Wilkinson RW, Cumberbatch M. Digital pattern recognition-based image analysis quantifies immune infiltrates in distinct tissue regions of colorectal cancer and identifies a metastatic phenotype. *Br J Cancer*. 2013;109(6):1618–1624.
16. Sinicrope FA, Mahoney MR, Smyrk TC, et al. Prognostic impact of deficient DNA mismatch repair in patients with stage III colon cancer from a randomized trial of FOLFOX-based adjuvant chemotherapy. *J Clin Oncol*. 2013;31(29):3664–3672.
17. Grambsch PM. Goodness-of-fit and diagnostics for proportional hazards regression models. *Cancer Treat Res*. 1995;75:95–112.
18. Harrell FJ. *Regression Modeling, Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2010.
19. Sargent DJ, Patiyl S, Yothers G, et al. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol*. 2007;25(29):4569–4574.
20. Sargent D, Shi Q, Yothers G, et al. Two or three year disease-free survival (DFS) as a primary end-point in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, C-06, C-07 and C89803. *Eur J Cancer*. 2011;47(7):990–996.
21. Galon J, Mlecnik B, Bindea G, et al. Towards the introduction of the “Immunoscore” in the classification of malignant tumours. *J Pathol*. 2014;232(2):199–209.
22. Grothey A, Sobrero AF, Shields AF, et al. Duration of adjuvant chemotherapy for stage III colon cancer. *N Engl J Med*. 2018;378(13):1177–1188.
23. Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature*. 2002;418(6901):934–934.
24. Galon J, Angell HK, Bedognetti D, Marincola FM. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity*. 2013;39(1):11–26.
25. Mlecnik B, Tosolini M, Charoentong P, et al. Biomolecular network reconstruction identifies T-cell homing factors associated with survival in colorectal cancer. *Gastroenterology*. 2010;138(4):1429–1440.
26. Hu W, Yang Y, Li X, et al. Multi-omics approach reveals distinct differences in left- and right-sided colon cancer. *Mol Cancer Res*. 2018;16(3):476–485.
27. Mima K, Sukawa Y, Nishihara R, et al. *Fusobacterium nucleatum* and T cells in colorectal carcinoma. *JAMA Oncol*. 2015;1(5):653–661.
28. Noshok K, Sukawa Y, Adachi Y, et al. Association of *Fusobacterium nucleatum* with immunity and molecular alterations in colorectal cancer. *World J Gastroenterol*. 2016;22(2):557–566.
29. Cercek A, Braghiroli MI, Chou JF, et al. Clinical features and outcomes of patients with colorectal cancers harboring NRAS mutations. *Clin Cancer Res*. 2017;23(16):4753–4760.
30. Vincent J, Mignot G, Chalmin F, et al. 5-fluorouracil selectively kills tumor-associated myeloid-derived suppressor cells resulting in enhanced T cell-dependent antitumor immunity. *Cancer Res*. 2010;70(8):3052–3061.