Novel Internationally Verified Method Reports Desmoplastic Reaction as the Most Significant Prognostic Feature For Disease-specific Survival in Stage II Colorectal Cancer

Ines P. Nearchou, BSc, Yoshiki Kajiwara, MD, PhD,† Satsuki Mochizuki, PhD,† David J. Harrison, MD,* Peter D. Caie, PhD,* and Hideki Ueno, MD, PhD†*

Abstract: Multiple histopathologic features have been reported as candidates for predicting aggressive stage II colorectal cancer (CRC). These include tumor budding (TB), poorly differentiated clusters (PDC), Crohn-like lymphoid reaction and desmoplastic reaction (DR) categorization. Although their individual prognostic significance has been established, their association with disease-specific survival (DSS) has not been compared in stage II CRC. This study aimed to evaluate and compare the prognostic value of the above features in a Japanese (n = 283) and a Scottish (n = 163) cohort, as well as to compare 2 different reporting methodologies: analyzing each feature from across every tissue slide from the whole tumor and a more efficient methodology reporting each feature from a single slide containing the deepest tumor invasion. In the Japanese cohort, there was an excellent agreement between the multi-slide and single-slide methodologies for TB, PDC, and DR ($\kappa = 0.798$ to 0.898) and a good agreement when assessing Crohn-like lymphoid reaction ($\kappa = 0.616$). TB (hazard ratio [HR] = 1.773; $P = 0.016$), PDC (HR = 1.706; $P = 0.028$), and DR (HR = 2.982; $P < 0.001$) based on the single-slide method were all significantly associated with DSS. DR was the only candidate feature reported to be a significant independent prognostic factor (HR = 2.982; $P < 0.001$) with both multi-slide and single-slide methods. The single-slide result was verified in the Scottish cohort, where multivariate Cox regression analysis reported that DR was the only significant independent feature (HR = 1.778; $P = 0.002$) associated with DSS. DR was shown to be the most significant of all the analyzed histopathologic features to predict disease-specific death in stage II CRC. We further show that analyzing the features from a single-slide containing the tumor's deepest invasion is an efficient and quicker method of evaluation.

Key Words: desmoplastic reaction, tumor budding, poorly differentiated clusters, prognosis, colorectal cancer

(*Am J Surg Pathol* 2019;00:000–000)

Colorectal cancer (CRC) is one of the most common cancers worldwide. To date, the tumor node metastasis staging is the basis for both prognostic stratification and therapeutic management of CRC patients.[1,2] However, a growing awareness exists that due to the heterogeneity present within the tumor microenvironment of CRC, additional prognostic features may prove valuable in subdividing CRC patients into more accurate prognostic groups and to improve their therapeutic consequences. This need is particularly relevant to stage II CRC patients where around 20% of the patients experience poor outcome.[3]

Over the past decades, extensive research has been conducted to identify significant prognostic features that improve upon tumor node metastasis prediction of stage II CRC patient outcome. Such include features of the tumor such as tumor budding (TB)[4–8] or poorly differentiated clusters (PDC)[9–14] as well as the host's interaction with the tumor such as the lymphocytic[15–18] and desmoplastic reaction (DR).[19–21]

Although previous studies have compared and established the prognostic value of TB and PDC in CRC,[4–6,8–14] their clinical assessment has not yet been implemented in practice, mainly due to the nonstandardized methodology for its assessment; something that the International Tumor Budding Consensus Conference aimed to address.[2]

The lymphocytic reaction within the tumor microenvironment has shown to be a prognostic factor in CRC.

From the *Department of Quantitative and Digital Pathology, School of Medicine, University of St Andrews, St Andrews, UK; and †Department of Surgery, National Defense Medical College, Saitama, Japan.
IMMUNOSCORE is a registered trademark owned by INSERM.
Conflicts of Interest and Source of Funding: This study was funded by Medical Research Scotland and the Japan Society for the Promotion of Science in collaboration with the British Council. The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.
Correspondence: Ines P. Nearchou, BSc, Department of Quantitative and Digital Pathology, School of Medicine, University of St Andrews, St Andrews KY16 9TF, UK (e-mail: i.p.nearchou@gmail.com).
Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.ajsp.com.

Previous research has demonstrated its reporting by the evaluation of nodular lymphoid aggregates referred to as Crohn-like lymphoid reaction (CLR).[17,22–24] Another factor concerning the host's reaction to the tumor is desmoplasia, defined by the presence of fibrosis at the tumor's invasive front.[19–21,25] Histologic classification of the DR in stage II, III, and IV CRC, has been shown to be an independent predictor of relapse-free survival and disease-specific survival (DSS).[21]

Unlike studies on TB and PDC, the prognostic value of CLR and DR has not been extensively verified, specifically interinstitutionally and internationally. The published methodology for evaluating the aforementioned features include the assessment of each diagnostic slide containing tumor and the identification of all those that include any tumor invasive front. CLR or DR are reported across each slide containing any invasive front, and TB or PDC "hotspots" are located before their quantification. Reporting each of the above features from the same single tissue section containing the deepest portion of the invasive front would be less labor-intensive and would reduce the time and effort required for their assessment by the pathologist. In this study, we firstly aimed to further verify and compare these feature's prognostic significance with DSS by applying the previously published guidelines for their assessment. Second, we aimed to explore how the reporting of these features across a single-slide would compare with their assessment across all slides containing any invasive front.

MATERIALS AND METHODS

Study Design and Patient Cohort

The patient material for this study comprised of an initial cohort of 283 stage II CRC patients who underwent surgical resection over the years 2006 to 2011 in the National Defense Medical College Hospital (NDMCH), Japan. A further international verification cohort of 163 patients treated between the years 2002 to 2005 from hospitals in Edinburgh, Scotland was studied. Neither cohort had been previously analyzed for the candidate prognostic features reported in this study. Associated clinicopathologic data included features such as pT stage, location of the tumor, differentiation, histologic type, and follow-up information; up to 9.3 years DSS for the Japanese cohort and up to 11.5 years for the Scottish cohort. Detailed patient characteristics can be found in Table 1 (Japanese cohort) and Supplementary Table 1 (Supplemental Digital Content 1, http://links.lww.com/PAS/A803; Scottish cohort). This study was approved by the Ethics Committee of the National Defense Medical College (approval ref: No.2992) and after review by the NHS Lothian NRS BioResource, REC-approved Research Tissue Bank (REC approval ref: 13/ES/0126), granted by East of Scotland Research Ethics Service. All samples had been anonymized before the commencement of the study.

Histopathologic Evaluation of Candidate Features

Hematoxylin and eosin (H&E) slides previously prepared for routine clinical reporting were used in this study. Slides from the NDMCH comprised specimens of longitudinal cut, whereas the ones from Edinburgh hospitals were of cross-sectional cut. All tumor slides were initially assessed to identify those containing any invasive front (multi-slide method). Once those were found, the features mentioned below were assessed using glass slides under an objective lens with ×20 magnification with a 20-mm eyepiece field number diameter (Olympus BX53, Tokyo, Japan) for the Japanese cohort. The median number of slides examined per patient was 6 (range, 1 to 21). The slide containing the deepest part of the invasive front of the tumor was then selected and revaluated (single-slide method). Findings from the single-slide methodology applied to the Japanese cohort were then verified on a single H&E slide, which contained the tumor's deepest invasion from the Scottish cohort. The evaluation of the single-slide from both the Japanese and Scottish cohorts was performed on digitized H&E slides. Whole slide H&E images were captured with a ×20 objective using a Leica Aperio AT2 (Leica Microsystems, Wetzlar, Germany) for the Japanese cohort and an Axioscan. Z1 (Zeiss, Oberkochen, Germany) for the Scottish cohort. Features were then assessed using the Aperio ImageScope pathology slide viewing software (version 12.3.3; Leica Microsystems), and the ZEN imaging software (version 2.3 blue edition; Zeiss) for each whole slide scanner's images respectively.

Tumor Budding

Tumor buds were defined as cancer clusters of up to 4 cells in the invasive front region. TB was initially assessed on the Japanese cohort according to the recommendations by the International Tumor Budding Consensus Conference.[2] Briefly, from all slides containing the invasive front, the slide containing the greatest degree of budding was firstly selected. Tumor buds were counted in the "hotspot" region at ×20 objective (area, 0.785 mm^2). Tumors with 0 to 4 buds per 0.785 mm^2 were classed as grade 1, 5 to 9 buds as grade 2 and $\geq$10 buds as grade 3 (Fig. 1A). Using the same categorization system, we reassessed TB on a single tissue slide containing the deepest part of the tumor on both the Japanese and Scottish cohorts.

Poorly Differentiated Clusters

Non–gland-forming cancer clusters of $\geq$5 cancer cells invading into the stroma were defined as PDC. Similarly to the TB assessment, all tissue slides containing the tumor front were initially assessed to identify the slide with the highest number of PDC. Within that slide, the region at ×20 objective where PDC was most intensively observed was assessed. Tumors with 0 to 4 PDC were classed as grade 1, 5 to 9 PDC as grade 2 and $\geq$10 PDC as grade 3 (Fig. 1B). The same categorization system was then used when applying the single-slide method.

Crohn-like Lymphoid Reaction

CLR assessment was performed as previously described.[17] Briefly, the presence of nodular lymphoid aggregates surrounding the tumor periphery of size $\geq$1 mm was classed as CLR (Fig. 1C). Nodular lymphoid aggregates of smaller size or present within or just below the mucosa layer were not evaluated as CLR and therefore

TABLE 1. Univariate and Multivariate Cox Regression Analysis For Clinicopathologic Data and Features Assessed on All Slides Containing the Invasive Front of the Japanese Cohort

		Univariate		Multivariate	
Features	**Frequency (%)**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Age		1.772 (0.944-3.325)	0.075	—	
≤70	177 (62.5)				
71-79	85 (30.0)				
≥80	21 (7.4)				
Sex		1.292 (0.560-2.979)	0.548	—	
Male	184 (65.0)				
Female	99 (35.0)				
pT stage		1.523 (0.612-3.794)	0.366	NS	
pT3	231 (81.6)				
pT4	52 (18.4)				
Tumor site		1.612 (0.959-2.709)	0.071	—	
Left	76 (26.9)				
Right	92 (32.5)				
Rectal	115 (40.6)				
Differentiation		1.047 (0.699-1.570)	0.822	NS	
Moderate	113 (39.9)				
Poor	25 (8.8)				
Well	145 (51.2)				
Tumor type				—	
Adenocarcinoma	266 (94.0)				
Mucinous	17 (6.0)				
DR on all slides		**2.980 (1.627-5.459)**	**< 0.001**	**2.980 (1.627-5.459)**	**< 0.001**
Mature	90 (31.8)				
Intermediate	105 (37.1)				
Immature	88 (31.1)				
PDC on all slides		1.613 (0.972-2.677)	0.065	NS	
G1	84 (29.7)				
G2	106 (37.5)				
G3	93 (32.9)				
TB on all slides		**1.749 (1.064-2.877)**	**0.028**	NS	
G1	103 (36.4)				
G2	84 (29.7)				
G3	96 (33.9)				
CLR on all slides		0.655 (0.296-1.447)	0.295	NS	
Present	146 (51.6)				
Absent	137 (48.4)				

Significant features ($P \leq 0.05$) are shown in bold.
G1 indicates grade 1; G2, grade 2; G3, grade 3; NS, not significant.

tumors with such features were classed as CLR absent tumors. CLR was initially assessed across all slides containing any invasive front in the Japanese cohort. CLR was then reassessed on a single-slide containing the tumor's deepest invasion from both the Japanese and Scottish cohorts.

Desmoplastic Reaction

DR was classified as immature, intermediate, or mature based on the 3-tier categorization system previously proposed.[21] Presence of myxoid stroma greater than a microscopic field of a ×40 objective lens within the extramural tumor front was regarded as immature (Fig. 1D). If myxoid stroma (greater than a microscopic field of a ×40 objective lens) was not present but the extramural tumor front included keloid-like collagens, the stroma was classified as intermediate (Fig. 1E). Absence of size significant myxoid stroma and keloid-like collagens was regarded as mature stroma (Fig. 1F). As per the evaluation of previous features, all slides containing any tumor invasive front from the Japanese cohort was initially assessed for DR before the slide containing only the deepest part of the tumor of the Japanese and Scottish cohorts was reassessed.

Statistical Analyses

To assess the agreement between the 2 methodologies used to evaluate the histopathologic features, the weighted κ was calculated.[26–28] The weighted κ was also calculated for assessing the agreement among the observers (I.P.N. and Y.K.) for the single-slide method of each feature reported on a randomly selected sample set (n = 50, 25 from the Scottish cohort and 25 from the Japanese cohort). Associations among the evaluated features and between the features and the data from the clinicopathologic report were analyzed using the χ^2 test. Univariate Cox regression was performed to assess the risk of disease-specific death over time for all histopathologic

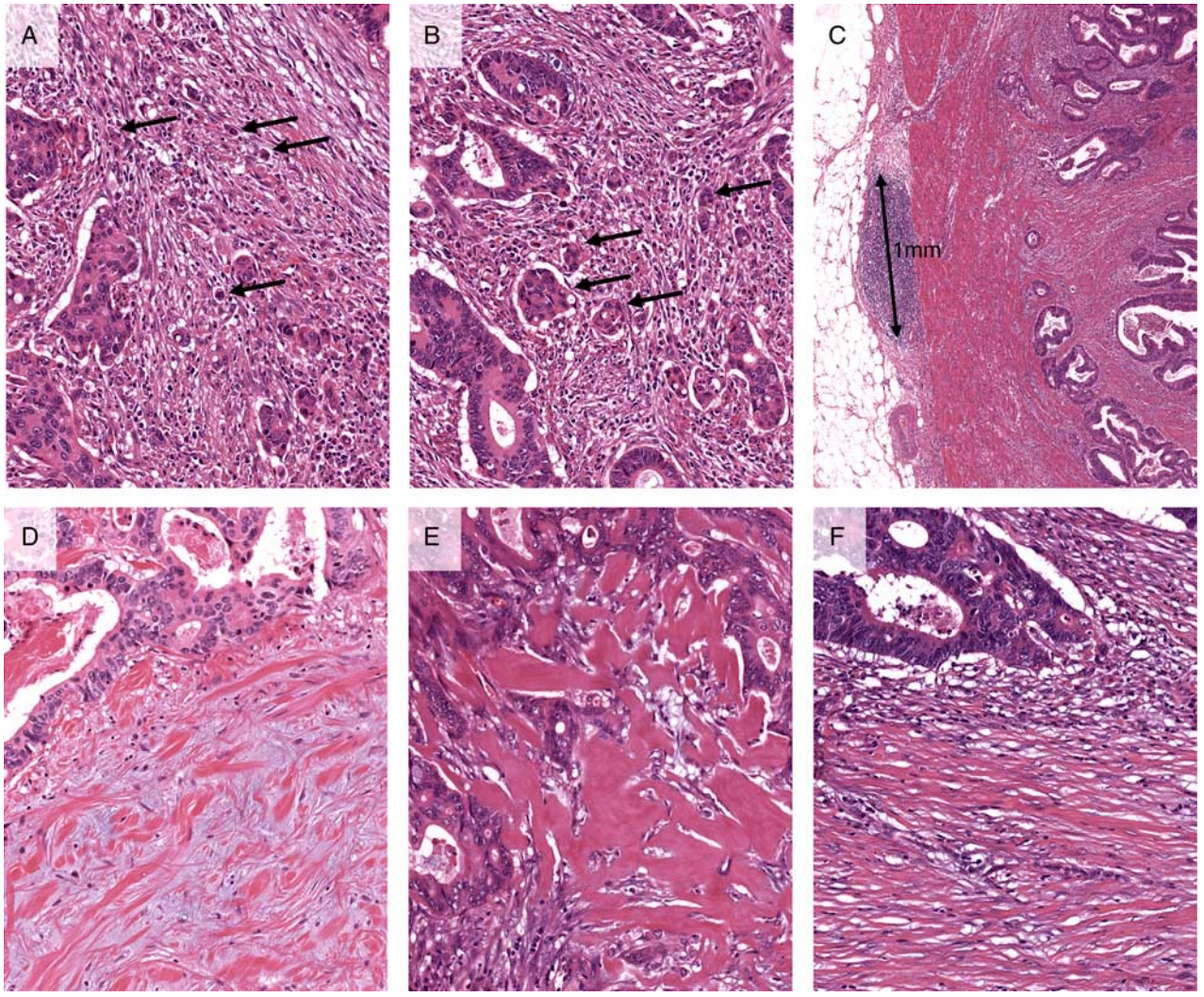

FIGURE 1. Histopathologic findings of tumor buds, PDC, CLR, and DR on H&E-stained slides. A, TB (cancer clusters of up to 4 cells) are shown by the black arrows. B, PDC (cancer clusters ≥ 5 cells) are shown by the black arrows. C, CLR present below the muscularis mucosa. D, Myxoid stroma defined by the presence of enlarged extracellular matrix and keloid like collagens with random directionality (immature type DR). E, Keloid-like collagen, that is, thick bands of collagen with bright eosinophilic hyalinization (intermediate type DR). F, Mature type DR with fibrotic stroma without myxoid stroma and keloid-like collagen.

features evaluated. Multivariate Cox regression with a forward stepwise method was applied in SPSS 24,[29] to select the most significant prognostic histopathologic features that added value to any final model. *P*-values ≤ 0.05 were regarded as statistically significant. Kaplan-Meier (KM) analysis was used to plot the survival curves of the 4 candidate histopathologic features assessed. Unless otherwise stated, R studio 1.1.419 running R 3.4.3[30,31] was used to perform all above analyses.

RESULTS

Patients' Characteristics

Patients' characteristics from both Japanese and Scottish cohorts are shown in Table 1 and Supplementary Table 1 (Supplemental Digital Content 1, http://links.lww.com/PAS/A803), respectively. This study included a Japanese cohort of 283 stage II CRC patients of which 99 were female and 184 male. Within the Scottish cohort (N = 163), 76 were female and 87 were male patients. Within the Japanese cohort, 231 patients were of pT3 and 52 of the pT4 stage, whereas there were 110 pT3 and 53 pT4 stage patients in the Scottish cohort. The age of patients ranged from 31 to 99 years (median 68 y) and from 37 to 96 years (median, 74 y) in the Japanese and Scottish cohorts respectively.

Method Agreement Analysis

Two separate quantification methodologies were assessed for the evaluation of the candidate prognostic features; TB, PDC, CLR, and DR. The first methodology was based on the previous recommendations for their reporting, where all glass tissue slides containing any invasive front are

assessed. The second methodology reports all candidate features from a single digitized tissue slide containing the deepest tumor invasion. To assess the agreement between the 2 methodologies, the weighted κ was calculated. Results revealed that the 2 methodologies were in excellent agreement for the assessment of TB, PDC, and DR (κ = 0.898, 0.798, and 0.827, respectively). The κ value for the agreement between the 2 methods for the evaluation for CLR was shown to be good (κ = 0.616) (Supplementary Table 2, Supplemental Digital Content 1, http://links.lww.com/PAS/A803).

Interobserver Agreement on Single-slide Method

Two observers (I.P.N. and Y.K.) scored a subset of 50 patients, 25 from each cohort, for all candidate histopathologic features. The κ value for the agreement of the assessments was excellent for TB, PDC, and DR (0.809, 0.801, 0.829, respectively), however, was moderate for CLR (0.448).

Correlations between Features

Features Assessed on Multiple Slides

The relationship between TB, PDC, CLR, and DR evaluated using the multi-slide method was assessed by the χ^2 test. Their relationship with the patient data was also assessed using the χ^2 test. DR was significantly associated with advanced T stage ($P = 0.003$). TB was associated with tumor type ($P = 0.038$) (Supplementary Table 3, Supplemental Digital Content 1, http://links.lww.com/PAS/A803). No other candidate feature was associated with any clinicopathologic characteristic. The degree of TB was significantly associated with the degree of PDC ($P < 0.001$) and DR ($P = 0.001$). PDC was also correlated with DR ($P < 0.001$). No association was found between CLR and any other candidate feature (Supplementary Table 4, Supplemental Digital Content 1, http://links.lww.com/PAS/A803).

Features Assessed on a Single-slide

The χ^2 test was further applied to analyze the associations between the candidate features, utilizing the single-slide methodology, and the data from the clinicopathologic report. Within the Japanese cohort, there was a statistically significant correlation between the PDC, CLR, DR, and pT stage ($P = 0.040$, 0.011, and 0.002, respectively) (Supplementary Table 5, Supplemental Digital Content 1, http://links.lww.com/PAS/A803). TB ($P < 0.001$), PDC ($P < 0.001$), and CLR ($P = 0.005$) were all significantly associated with DR. TB was also significantly associated with CLR ($P = 0.024$) and PDC ($P < 0.001$) (Supplementary Table 6, Supplemental Digital Content 1, http://links.lww.com/PAS/A803). Results from the Scottish cohort showed that PDC and TB were both significantly associated with pT stage ($P = 0.006$ and <0.001, respectively) and differentiation ($P = 0.016$ and <0.001, respectively) (Supplementary Table 7, Supplemental Digital Content 1, http://links.lww.com/PAS/A803). Furthermore, TB was shown to be significantly correlated with the degree of PDC ($P < 0.001$) as well as DR ($P < 0.001$). PDC was also shown to be associated with DR ($P = 0.010$). CLR was not shown to be significantly associated with any other candidate feature in the Scottish cohort (Supplementary Table 8, Supplemental Digital Content 1, http://links.lww.com/PAS/A803).

Survival Analysis

Features Assessed on Multiple Slides From the Japanese Cohort

Univariate Cox regression was performed to assess the prognostic significance of all clinicopathologic data and the candidate features assessed on all slides containing the invasive front. None of the clinicopathologic data, including pT stage and differentiation, were reported as significant predictors of DSS. Of the candidate histopathologic features assessed in this study, TB (hazard ratio [HR] = 1.749; 95% confidence interval [CI], 1.064-2.877; $P = 0.028$) and DR (HR = 2.980; 95% CI, 1.627-5.459; $P < 0.001$) were significantly associated with DSS, whereas PDC only trended to significance. CLR was not reported as significant in this study (Table 1). KM survival analysis was calculated for the 4 candidate features; TB, PDC, CLR, and DR (Fig. 2). Results showed that in regard to TB, the 9.3-year DSS was 87.4% for grade 1, 69.4% for grade 2, and 63.9% for grade 3. This trend was similarly observed when assessing the DSS rate for PDC; 88.1% for grade 1, 77.0% for grade 2, and 60.3% for grade 3. No discernable difference was observed in the survival rates according to CLR (73.3% survival for negative and 72.8% survival for positive). The 9.3-year DSS for the DR categorization was reported as 97.5% for the mature group, 74.8% for the intermediate group, and only 50.3% in the immature. DR ($P = 0.0003$) was the only candidate feature to significantly stratify the patients within this cohort by KM analysis. When assessing which features added value to the prediction of disease-specific death, forward stepwise multivariate Cox regression was applied and the predictive model was adjusted for TB, PDC, CLR, DR, and the gold standards of pT stage and differentiation. Results showed only DR (HR = 2.980; 95% CI, 1.627-5.459; $P < 0.001$) to be an independent significant survival predictor (Table 1).

Features Assessed on a Single-slide of the Japanese Cohort

Following the reassessment of all candidate features reported from the same single-slide that contained the deepest invasion of the tumor, univariate Cox regression was applied to assess their prognostic significance (Table 2). The results revealed that TB (HR = 1.773; 95% CI, 1.111-2.829; $P = 0.016$), PDC (HR = 1.706; 95% CI, 1.061-2.745; $P = 0.028$) and DR (HR = 2.982; 95% CI, 1.702-5.223; $P < 0.001$) were all significantly associated with a risk of disease-specific death. CLR was again not shown to have a prognostic significance. When KM survival curves were plotted, the results reflected those of the multi-slide methodology, where patients were stratified based on the reporting of the candidate features, albeit only significantly for TB ($P = 0.047$) and DR ($P < 0.0001$) (Fig. 2). Multivariate forward stepwise Cox regression analysis adjusted for TB, PDC, CLR, and DR, resulted in DR to

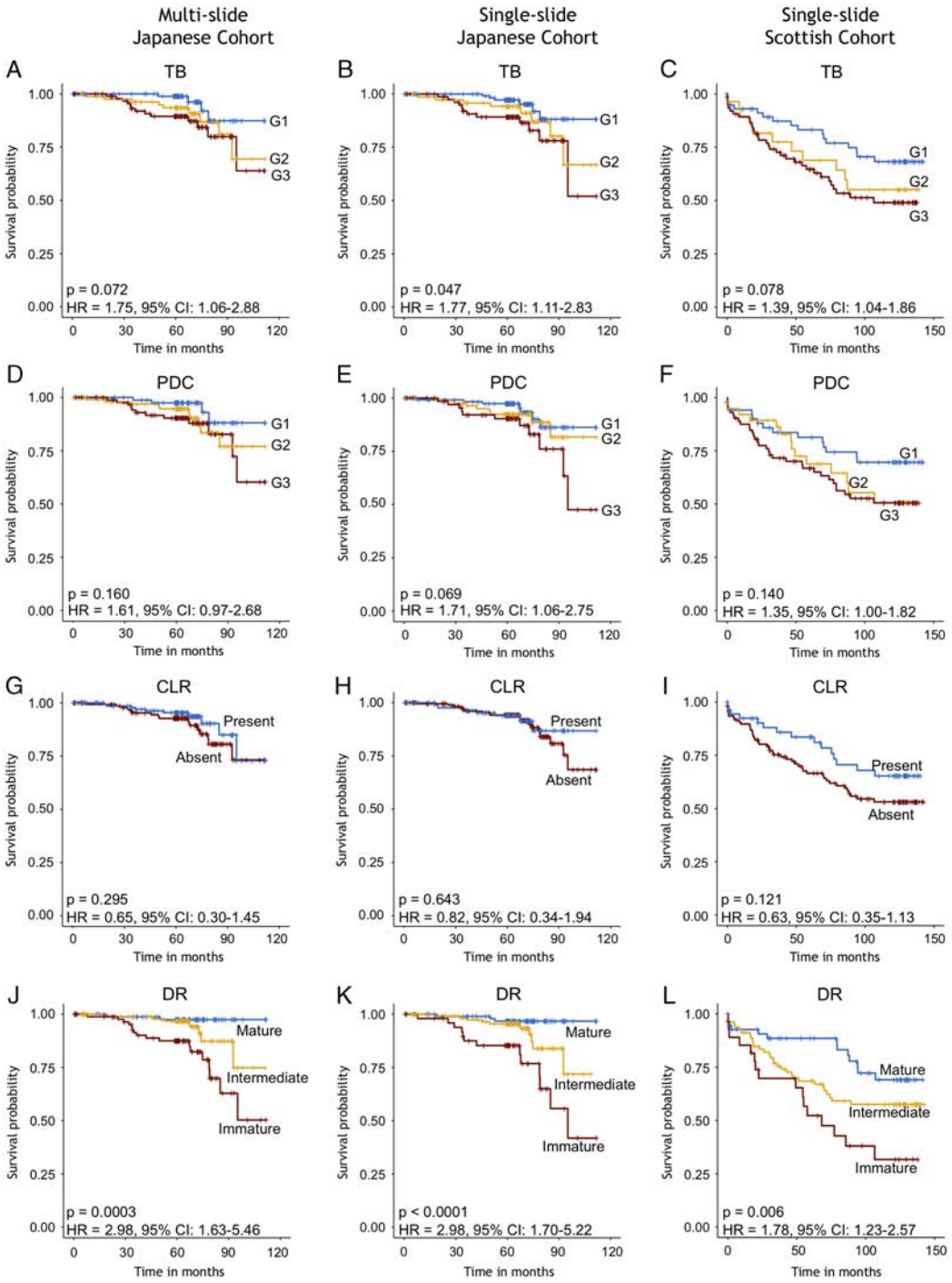

FIGURE 2. KM survival analysis for TB grade, PDC grade, CLR, and categorization of DR for the Japanese cohort (using all slides containing any invasive front and a single-slide containing the deepest tumor invasion) and for the Scottish cohort (single-slide). A, TB evaluated on all tissue slides containing any invasive front from the Japanese cohort. B, TB assessed on a single tissue slide containing the deepest tumor invasion from the Japanese cohort. C, TB assessed on a single tissue slide containing the deepest tumor invasion from the Scottish cohort. D, PDC evaluated on all tissue slides containing any invasive front from the Japanese cohort. E, PDC assessed on a single tissue slide containing the deepest tumor invasion from the Japanese cohort. F, PDC assessed on a single tissue slide containing the deepest tumor invasion from the Scottish cohort. G, CLR evaluated on all tissue slides containing any invasive front from the Japanese cohort. H, CLR assessed on a single tissue slide containing the deepest tumor invasion from the Japanese cohort. I, CLR assessed on a single tissue slide containing the deepest tumor invasion from the Scottish cohort. J, DR evaluated on all tissue slides containing any invasive front from the Japanese cohort. K, DR assessed on a single tissue slide containing the deepest tumor invasion from the Japanese cohort. L, DR assessed on a single tissue slide containing the deepest tumor invasion from the Scottish cohort. *P* values from KM analysis. HR and CI from univariate Cox regression analysis.

TABLE 2. Univariate and Multivariate Cox Regression Analysis For Features Assessed on a Single Tissue Slide of the Japanese and Scottish Cohorts

		Univariate		Multivariate	
Features	**Frequency (%)**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Japanese cohort					
DR on 1 slide		**2.982 (1.702-5.223)**	**< 0.001**	**2.982 (1.702-5.223)**	**< 0.001**
Mature	107 (37.8)				
Intermediate	125 (44.2)				
Immature	51 (18.0)				
PDC on 1 slide		**1.706 (1.061-2.745)**	**0.028**	NS	
G1	124 (43.8)				
G2	88 (31.1)				
G3	71 (25.1)				
TB on 1 slide		**1.773 (1.111-2.829)**	**0.016**	NS	
G1	124 (43.8)				
G2	77 (27.2)				
G3	82 (29.0)				
CLR on 1 slide		0.815 (0.342-1.940)	0.643	NS	
Present	91 (32.2)				
Absent	192 (67.8)				
Scottish cohort					
DR on 1 slide					
Mature	56 (34.4)	**1.778 (1.230-2.571)**	**0.002**	**1.779 (1.230-2.571)**	**0.002**
Intermediate	79 (48.5)				
Immature	28 (17.2)				
PDC on 1 slide		1.349 (0.997-1.824)	0.052	NS	
G1	52 (31.9)				
G2	39 (23.9)				
G3	72 (44.2)				
TB on 1 slide		**1.391 (1.038-1.863)**	**0.027**	NS	
G1	59 (36.2)				
G2	28 (17.2)				
G3	76 (46.6)				
CLR on 1 slide		0.629 (0.351-1.129)	0.121	NS	
Present	55 (66.3)				
Absent	108 (33.7)				

Significant features ($P \leq 0.05$) are shown in bold.
NS indicates not significant.

be the only significant predictor of survival (HR = 2.982; 95% CI, 1.702-5.223; $P < 0.001$) (Table 2).

Single-slide Method Verified on Scottish Cohort

The faster and more efficient single-slide methodology was tested, for verification purposes, on an internationally distinct cohort of Scottish stage II CRC patients. Univariate Cox regression was applied to test the prognostic significance of TB, PDC, CLR, and DR (Table 2) as well as the data from the clinicopathologic report (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/PAS/A803). The results showed that TB (HR = 1.391; 95% CI, 1.038-1.863; $P = 0.027$), DR (HR = 1.778; 95% CI, 1.230-2.571; $P = 0.002$), pT stage (HR = 4.001; 95% CI, 2.384-6.715; $P < 0.001$), and differentiation (HR = 1.481; 95% CI, 1.235-1.775; $P < 0.001$) were statistically significant, although PDC did trend to significance. As with the methodology applied to the Japanese cohort, CLR was not reported as significant when assessing DSS. KM analysis was calculated to plot the patient stratification and DSS rates for TB, PDC, CLR, and DR. Higher grades of TB or PDC and the absence of CLR was associated with shorted DSS. The reporting of DR was the only feature to significantly stratify the Scottish patients ($P = 0.006$), where immature stroma was shown to be associated with the shortest DSS. CLR was again unable to significantly stratify the patients according to the DSS (Fig. 2). Multivariate forward stepwise Cox regression analysis with the candidate features of TB, PDC, CLR, and DR as input to the model, showed that as with the Japanese cohort, DR was the only independent predictor of shorter DSS (HR = 1.779; 95% CI, 1.230-2.572; $P = 0.002$) (Table 2).

DISCUSSION

The observations from this study showed that DR was the strongest prognostic indicator of DSS in stage II CRC when compared with TB, PDC, and CLR. DR was the only feature that added value to a multivariate Cox regression model in both interinstitutional cohorts and when applying the multi-slide or the single-slide analysis. In contrast, the current clinical gold standard for CRC patient stratification, pT stage, was only significantly correlated with DSS in the Scottish cohort. This finding, therefore, suggests that although pT stage is an established predictor of survival in stage II CRC, DR evaluation

might improve this staging system's accuracy in predicting DSS.

This study not only examined and compared the prognostic significance of known histopathologic markers in CRC through their recommended assessment methodologies but also through the assessment across only 1 tissue slide containing the deepest advancing edge of the tumor. This single-slide method brings advantages such as conserving the time and effort required for the assessment of these features while allowing their accurate evaluation with respect to patient prognosis. Furthermore, this novel methodology was verified across 2 distinct and international stage II CRC cohorts. The agreement between the multi-slide and single-slide methods of analysis was excellent for TB, PDC, and DR and good for CLR; suggesting that the assessment of only 1 tissue slide, containing the deepest invasion of the tumor, is potentially sufficient to stratify patients into prognostic categories, specifically for those features that were significantly associated with DSS. The interobserver agreement for TB, PDC, and DR were all excellent; however, the reporting of DR had the highest κ coefficient. This demonstrates that the novel single-slide method for assessing TB, PDC, and DR reported here, is reproducible and could be a viable and faster method for the evaluation of these features in the future.

Both TB and PDC represent the aggressive invasion of the tumor through the form of dedifferentiation and both have previously been linked to adverse clinical outcome in patients with CRC.[10,32–34] Our results showed that TB was significantly associated with DSS regardless of their assessment methodology, whereas PDC was only significant when assessed using the single-slide method, even though it did trend to significance using the recommended multi-slide assessment methodology. In regard to TB, this finding concurs well with previous results reported in the literature, whereas regarding PDC, these results differ from those shown in previous studies.[9–11] However, our results from both methodologies and cohorts revealed a strong correlation between TB and PDC grade. This agrees with the work previously published by Konishi et al[35] while supports the idea of a possible combined PDC and TB grade as a dedifferentiation marker in stage II CRC, recently shown by Lee and Chan.[36]

In an attempt to standardize CLR reporting, we have previously proposed a semiquantitative methodology for the CLR assessment based on the presence or absence of lymphoid aggregates of size $\geq$1mm.[17] However, in this study using the single-slide method, CLR still had a moderate interobserver concordance and held no significant prognostic value. This is in good agreement with a recent study,[35] where CLR was shown to have the poorest prognostic accuracy of relapse-free survival in CRC among other factors of the tumor invasive front, including the ones assessed in this study and a moderate interobserver agreement. Although assessment of CLR is readily available using routine H&E reporting, the IMMUNOSCORE, while reliant on immunohistochemistry and specific commercial image analysis software, has been demonstrated to be reproducible and significant in stage II CRC.[16,37] The IMMUNOSCORE measures specific densities of individual lymphocytes within the tumor microenvironment and so may be a more accurate representation of lymphocytic infiltration. This, therefore, indicates that methodologies like IMMUNOSCORE, even though more resource intensive, show better potential for clinical translation of the assessment of the host lymphocytic reaction.

Desmoplasia, defined by the presence of fibrosis at the tumor's invasive front, has been shown to be associated with cancer aggressiveness in a number of tumor types.[19,25,38–42] Specifically, in CRC, DR has been previously shown to have a significant association with T stage, lymphatic and venous invasion.[19] Previous research by our team has suggested a histologic classification of the DR into 3 prognostic categories; mature, intermediate, or immature, based on the keloid-like collagen presence and myxoid stroma at the extramural desmoplastic front.[21,25] These studies have shown the survival outcome to be most favorable in patients with mature stroma, followed by the ones with intermediate stroma and poorest in patients with immature stroma. Similar to our previous studies, we found DR to have a high prognostic significance using the same categorization system. Interestingly, DR was also significant when assessed through our novel method and after international verification, strongly supporting the promising role of DR assessment in the stratification of stage II patients into prognostic subgroups based on DSS. This fits well with the study by Konishi et al,[35] where although PDC was shown to be the most significant prognostic marker for colon cancer recurrence, DR was reported to be a significant prognostic factor in their large single-institute study. In our study, DR was also shown to be significantly associated with the degree of TB or PDC regardless of the methodology of assessment or cohort. These results confirm our previous findings where a higher degree of TB was observed more frequently in immature stroma[20,21] as well as the recent work by Konishi et al,[35] where the degree of PDC was significantly correlated with DR. Further, it strongly highlights the pivotal role the stroma has on tumor aggressiveness and its potential to metastasize via tumor de-differentiation.

A key strength of the present study was the evaluation and verification of the novel methodology on 2 international cohorts which consisted of tumor tissues sectioned differently (longitudinal vs. cross-section). Furthermore, as pathology moves closer to the full adoption of digital imaging, the single-slide analysis was performed using digitized whole slide images of the glass H&E in both the Japanese and Scottish cohorts, thus allowing the methodology to be future proofed. The glass reporting of the multi-slide method showed excellent agreement with the single-slide digital analysis for the reporting of DR. Furthermore, digital pathology allowed for rapid collaboration between multidisciplinary and international colleagues, as geographic barriers could be broken down. Finally, to make the analysis as democratic as possible, as well as scanner-agnostic, the digitization of the slides was performed using 2 separate scanning platforms and slide viewing software, one for each of the international cohorts. Despite the presence of these differences, the results reporting DR as the most significant prognostic feature for DSS were

consistent across methods, international cohorts and independent of scanner type.

A limitation of this study is the absence of data on microsatellite status for both cohorts, hence their incorporation into the statistical analysis was not possible. In addition, the cases presented in this study were of the most common histologic types (adenocarcinoma, mucinous, or mixed), therefore it would be interesting to study the prognostic significance of DR in more rare and aggressive subtypes of stage II CRC.

We present the first study to examine and compare TB, PDC, CLR, and DR with relation to stage II CRC DSS and their reporting from a single tissue slide, rather than the evaluation of multiple slides. Furthermore, we are the first to directly verify these findings in an interinstitutional and international study. Through both the recommended methodologies and the newly proposed efficient method, we have shown that DR was the most significant prognostic feature when compared with TB, PDC, and CLR. The present findings not only suggest that DR categorization is a promising prognostic factor for stage II CRC but also that the fibrotic stroma at the invasive front of the tumor may play a fundamental role in determining the CRC biological behavior. However, further studies concerning the molecular background of DR would enable the better understanding of the controlling influence the desmoplastic fibrotic stroma has on the tumor itself.

ACKNOWLEDGMENTS

The authors thank Frances Rae from Tissue Governance NHS Lothian for securing ethics and patient follow-up data as well as Dr Kengo Takeuchi and Dr Naoko Tsuyama for kindly allowing the digitization of the Japanese slides to be performed at the Cancer Institute Hospital of Japanese Foundation for Cancer Research. They also thank Medical Research Scotland, the Japan Society for the Promotion of Science and the British Council for funding this study.

REFERENCES

1. Li J, Yi C-H, Hu Y-T, et al. TNM staging of colorectal cancer should be reconsidered according to weighting of the T stage: verification based on a 25-year follow-up. *Medicine (Baltimore)*. 2016;95:e2711.
2. Lugli A, Kirsch R, Ajioka Y, et al. Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Mod Pathol*. 2017;30:1299–1311.
3. Nitsche U, Maak M, Schuster T, et al. Prediction of prognosis is not improved by the seventh and latest edition of the TNM classification for colorectal cancer in a single-center collective. *Ann Surg*. 2011;254:793–801.
4. Ueno H, Murphy J, Jass JR, et al. Tumour "budding" as an index to estimate the potential of aggressiveness in rectal cancer. *Histopathology*. 2002;40:127–132.
5. Tanaka M, Hashiguchi Y, Ueno H, et al. Tumor budding at the invasive margin can predict patients at high risk of recurrence after curative surgery for stage II, T3 colon cancer. *Dis Colon Rectum*. 2003;46:1054–1059.
6. Ueno H, Price AB, Wilkinson KH, et al. A new prognostic staging system for rectal cancer. *Ann Surg*. 2004;240:832–839.
7. Horcic M, Koelzer VH, Karamitopoulou E, et al. Tumour budding score based on 10 high-power fields is a promising basis for a standardised prognostic scoring system in stage II colorectal cancer. *Hum Pathol*. 2013;44:697–705.
8. Caie PD, Turnbull AK, Farrington SM, et al. Quantification of tumour budding, lymphatic vessel density and invasion through image analysis in colorectal cancer. *J Transl Med*. 2014;12:156.
9. Ueno H, Mochizuki H, Hashiguchi Y, et al. Histological grading of colorectal cancer: a simple and objective method. *Ann Surg*. 2008;247:811–818.
10. Ueno H, Hase K, Hashiguchi Y, et al. Site-specific tumor grading system in colorectal cancer: multicenter pathologic review of the value of quantifying poorly differentiated clusters. *Am J Surg Pathol*. 2014;38:197–204.
11. Ueno H, Kajiwara Y, Shimazaki H, et al. New criteria for histologic grading of colorectal cancer. *Am J Surg Pathol*. 2012;36:193–201.
12. Caie PD, Zhou Y, Turnbull AK, et al. Novel histopathologic feature identified through image analysis augments stage II colorectal cancer clinical reporting. *Oncotarget*. 2016;7:44381–44394.
13. Bonetti LR, Barresi V, Bettelli S, et al. Poorly differentiated clusters (PDC) in colorectal cancer: what is and ought to be known. *Diagn Pathol*. 2016;11:31.
14. Barresi V, Reggiani Bonetti L, Ieni A, et al. Poorly differentiated clusters: clinical impact in colorectal cancer. *Clin Colorectal Cancer*. 2017;16:9–15.
15. Pagès F, Kirilovsky A, Mlecnik B, et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol*. 2009;27:5944–5951.
16. Pagès 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;391:2128–2139.
17. Ueno H, Hashiguchi Y, Shimazaki H, et al. Objective criteria for Crohn-like lymphoid reaction in colorectal cancer. *Am J Clin Pathol*. 2013;139:434–441.
18. Lugli A, Karamitopoulou E, Panayiotides I, et al. CD8+ lymphocytes/tumour-budding index: an independent prognostic factor representing a "pro-/anti-tumour" approach to tumour host interaction in colorectal cancer. *Br J Cancer*. 2009;101:1382–1392.
19. Ueno H, Shinto E, Shimazaki H, et al. Histologic categorization of desmoplastic reaction: its relevance to the colorectal cancer microenvironment and prognosis. *Ann Surg Oncol*. 2015;22:1504–1512.
20. Ueno H, Jones AM, Wilkinson KH, et al. Histological categorisation of fibrotic cancer stroma in advanced rectal cancer. *Gut*. 2004;53:581–586.
21. Ueno H, Kanemitsu Y, Sekine S, et al. Desmoplastic pattern at the tumor front defines poor-prognosis subtypes of colorectal Cancer. *Am J Surg Pathol*. 2017;41:1506–1512.
22. Kim JH, Kim K-J, Bae JM, et al. Comparative validation of assessment criteria for Crohn-like lymphoid reaction in colorectal carcinoma. *J Clin Pathol*. 2015;68:22–28.
23. Rozek LS, Schmit SL, Greenson JK, et al. Tumor-infiltrating lymphocytes, Crohn's-like lymphoid reaction, and survival from colorectal cancer. *J Natl Cancer Inst*. 2016:108.
24. Ogino S, Nosho K, Irahara N, et al. Lymphocytic reaction to colorectal cancer is associated with longer survival, independent of lymph node count, MSI and CpG island methylator phenotype. *Clin Cancer Res*. 2009;15:6412–6420.
25. Ueno H, Sekine S, Oshiro T, et al. Disentangling the prognostic heterogeneity of stage III colorectal cancer through histologic stromal categorization. *Surgery*. 2018;163:777–783.
26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
27. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213–220.
28. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychol Bull*. 1969;72:323–327.
29. IBM Corp. *IBM SPSS Statistics for Windows, Version 240*. Armonk, NY: IBM Corp.; 2016.
30. RStudio Team. *RStudio: Integrated Development for R*. Boston, MA: RStudio Inc.; 2015.
31. R Core Team. *R: The R Project for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available at: https://www.r-project.org/. Accessed October 22, 2018.

32. Zlobec I, Lugli A, Baker K, et al. Role of APAF-1, E-cadherin and peritumoural lymphocytic infiltration in tumour budding in colorectal cancer. *J Pathol*. 2007;212:260–268.
33. Kajiwara Y, Ueno H, Hashiguchi Y, et al. Heterogeneity of metalloproteinase expression in colorectal cancer—relation of molecular findings to basic morphology. *Anticancer Res*. 2011;31:1567–1575.
34. Kajiwara Y, Ueno H, Hashiguchi Y, et al. Expression of L1 cell adhesion molecule and morphologic features at the invasive front of colorectal cancer. *Am J Clin Pathol*. 2011;136:138–144.
35. Konishi T, Shimada Y, Lee LH, et al. Poorly differentiated clusters predict colon cancer recurrence. *Am J Surg Pathol*. 2018;42:705–714.
36. Lee VWK, Chan KF. Tumor budding and poorly-differentiated cluster in prognostication in stage II colon cancer. *Pathol Res Pract*. 2018;214:402–407.
37. Galon J, Mlecnik B, Bindea G, et al. Towards the introduction of the "Immunoscore" in the classification of malignant tumours. *J Pathol*. 2014;232:199–209.
38. Lagacé R, Grimaud J-A, Schürch W, et al. Myofibroblastic stromal reaction in carcinoma of the breast: variations of collagenous matrix and structural glycoproteins. *Virchows Arch A Pathol Anat Histopathol*. 1985;408:49–59.
39. Shimosato Y, Suzuki A, Hashimoto T, et al. Prognostic implications of fibrotic focus (scar) in small peripheral lung cancers. *Am J Surg Pathol*. 1980;4:365–373.
40. Zippi M, De Toma G, Minervini G, et al. Desmoplasia influenced recurrence of disease and mortality in stage III colorectal cancer within five years after surgery and adjuvant therapy. *Saudi J Gastroenterol*. 2017;23:39–44.
41. Wang LM, Silva MA, D'Costa Z, et al. The prognostic role of desmoplastic stroma in pancreatic ductal adenocarcinoma. *Oncotarget*. 2016;7:4183–4194.
42. Cardone A, Tolino A, Zarcone R, et al. Prognostic value of desmoplastic reaction and lymphocytic infiltration in the management of breast cancer. *Panminerva Med*. 1997;39:174–177.