ORIGINAL ARTICLE

Association of the collagen signature with pathological complete response in rectal cancer patients

Wei Jiang[1,2] | Shijie Wang[1] | Jinliang Wan[1] | Jixiang Zheng[1] | Xiaoyu Dong[1] | Zhangyuanzhu Liu[3] | Guangxing Wang[2] | Shuoyu Xu[1,4] | Weiwei Xiao[5] | Yuanhong Gao[5] | Shuangmu Zhuo[2] | Jun Yan[1]

[1]Department of General Surgery, Guangdong Provincial Key Laboratory of Precision Medicine for Gastrointestinal Tumor, Nanfang Hospital, The First School of Clinical Medicine, Southern Medical University, Guangzhou, China

[2]School of Science, Jimei University, Xiamen, China

[3]Department of Hepatobiliary and Pancreatic Surgery, Guangdong Provincial Hospital of Traditional Chinese Medicine, The Second Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine, Guangzhou, China

[4]Department of Radiology, Sun Yat-sen University Cancer Center, Guangzhou, China

[5]Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Correspondence

Jun Yan, Department of General Surgery, Guangdong Provincial Key Laboratory of Precision Medicine for Gastrointestinal Tumor, Nanfang Hospital, The First School of Clinical Medicine, Southern Medical University, Guangzhou, Guangdong, 510515, China.
Email: yanjunfudan@163.com

Shuangmu Zhuo, School of Science, Jimei University, Xiamen, Fujian, 361021, China.
Email: shuangmuzhuo@gmail.com

Yuanhong Gao, Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China.
Email: gaoyh@sysucc.org.cn

Funding information

This work was supported by grants from the National Natural Science Foundation of China (81773117 and 81771881), the State's Key Project of Research and Development Plan (2019YFE0113700),

Abstract

Collagen in the tumor microenvironment is recognized as a potential biomarker for predicting treatment response. This study investigated whether the collagen features are associated with pathological complete response (pCR) in locally advanced rectal cancer (LARC) patients receiving neoadjuvant chemoradiotherapy (nCRT) and develop and validate a prediction model for individualized prediction of pCR. The prediction model was developed in a primary cohort (353 consecutive patients). In total, 142 collagen features were extracted from the multiphoton image of pretreatment biopsy, and the least absolute shrinkage and selection operator (Lasso) regression was applied for feature selection and collagen signature building. A nomogram was developed using multivariable analysis. The performance of the nomogram was assessed with respect to its discrimination, calibration, and clinical utility. An independent cohort (163 consecutive patients) was used to validate the model. The collagen signature comprised four collagen features significantly associated with pCR both in the primary and validation cohorts ($p < 0.001$). Predictors in the individualized prediction nomogram included the collagen signature and clinicopathological predictors. The nomogram showed good discrimination with area under the ROC curve (AUC)

of 0.891 in the primary cohort and good calibration. Application of the nomogram in the validation cohort still gave good discrimination (AUC = 0.908) and good calibration. Decision curve analysis demonstrated that the nomogram was clinically useful. In conclusion, the collagen signature in the tumor microenvironment of pretreatment biopsy is significantly associated with pCR. The nomogram based on the collagen signature and clinicopathological predictors could be used for individualized prediction of pCR in LARC patients before nCRT.

KEYWORDS
collagen signature, neoadjuvant chemoradiotherapy, pathological complete response, rectal cancer, tumor microenvironment

1 | INTRODUCTION

Currently, colorectal cancer is one of the tumors with the highest morbidity and mortality, and LARC accounts for approximately 70% of rectal cancers.[1] To improve the rates of R0 resection and sphincter-preserving surgery, neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision is the standard treatment for LARC patients.[2] Approximately 20%–25% of patients achieve pathological complete response (pCR) after nCRT, and these patients experience a better prognosis than patients with non-pCR.[3]

Some researchers consider the proportion of patients who achieve pCR after nCRT and attempt to identify alternative treatment options for TME due to the surgery-related deaths and postoperative functional complications associated with TME, especially abdominoperineal resection. Habr-Gama et al.[4] found that the "wait and see" policy in clinical complete response patients compared with pCR patients who underwent TME showed no differences in prognosis. This original finding was subsequently supported by a series of studies[5,6]; therefore, the "wait and see" policy can be considered an alternative treatment strategy for TME. There is a significant clinical need for a reliable biomarker to accurately predict pCR in LARC patients who may safely adopt the "wait and see" policy after nCRT.

The tumor microenvironment consists of various tumor cell components and noncellular extracellular matrix (ECM), and the ECM interaction with tumor cells plays a critical role in tumor progression, metastasis, and therapeutic efficacy.[7,8] Collagen is the dominant component of the ECM, and its structure has been increasingly recognized as a robust biomarker to predict the prognosis of multiple tumor types, such as prostate cancer and gastric cancer.[9,10] Furthermore, previous studies found that the collagen structure in a biopsy is associated with the treatment response associated with nCRT in rectal cancer and breast cancer.[11,12] Nevertheless, the relationship between the collagen structure in pretreatment biopsy and pCR has not been examined. Therefore, we hypothesized that collagen structure in the tumor microenvironment of the biopsy is associated with pCR in LARC patients.

Multiphoton imaging (MPI) is a fast, label-free, high-resolution imaging technology that combines two nonlinear optical effects, the SHG signal generated by collagen and the two-photon excitation fluorescence (TPEF) signal for cells, to observe detailed information on collagen structure and cell morphology at the subcellular level.[13] Moreover, due to the inherent physical features of collagen, MPI has become a useful optical tool to visualize the collagen structure in the tumor microenvironment.[14,15] In addition, high-throughput and fully quantified collagen structure features can be extracted from high-resolution multiphoton images through the automatic image analysis method,[10,14,15] which aids in interpreting the relationship between collagen structure and pCR.

Here, we clarified the correlation between the collagen structure in the biopsy tumor microenvironment and pCR and then developed and validated a nomogram for accurately individualized prediction of pCR in patients with LARC before nCRT.

2 | MATERIAL AND METHODS

2.1 | Data collection

This study was approved by the Institutional Review Board at Nanfang Hospital, Sun Yat-sen University Cancer Center, and Fujian Province Tumor Hospital (approval number: NFEC-2021-440). Written informed consent was obtained from all participants. The study was conducted in compliance with the Declaration of Helsinki. The inclusion criteria were as follows: (i) LARC (≥T3 and/or N+) was diagnosed by pretreatment medical imaging and pathological examination; (ii) availability of pretreatment biopsy tissue; and (iii) standardized nCRT completed followed by surgical resection. The exclusion criteria were as follows: (i) short course radiotherapy and (ii) no radical surgery after nCRT. Based on the inclusion and exclusion criteria, we enrolled two independent panels of LARC patients with nCRT, including a primary cohort between January 2010 and June 2018. An independent validation cohort was obtained from January 2010 to June 2018.

Pretreatment tumor biopsy tissues were obtained endoscopically before nCRT, and two experienced gastrointestinal pathologists reassessed all of these biopsies. The pretreatment biopsy was fixed in formalin, embedded in paraffin, and cut into 4-μm sections for MPI.

The clinicopathologic characteristics collected from three medical records were as follows: age, sex, body mass index (BMI), differentiation status, pretreatment carcinoembryonic antigen (CEA) level, pretreatment carbohydrate antigen 199 (CA-199) level, distance from anal verge, pretreatment T stage, pretreatment N stage, and tumor dimension.

In this study, pretreatment biopsies were reviewed for evidence of tumor budding through a ×4 lens (×40 magnification) with confirmation of positive cases at ×10 (×100 magnification). Tumor budding was defined as a single cancer cell or a group of <5 detached tumor cells found in the stroma of the biopsy specimen.[16] Therefore, any budding seen at ×4 and confirmed at ×10 was deemed positive.

2.2 | Treatment and definition of pCR

All patients underwent preoperative radiotherapy at a total dose of 50.4 Gy in 28 fractions. Concomitantly, preoperative chemotherapy was delivered, and radiotherapy was administered according to National Comprehensive Cancer Network (NCCN) guidelines.[17] TME was performed within 6–8 weeks after completion of nCRT by senior attending surgeons. Adjuvant chemotherapy started within 6 weeks after surgery. The regimen was the same as that for preoperative chemotherapy.

The treatment response was evaluated by two gastrointestinal pathologists who were blind to the clinical outcomes according to surgical resection specimens. Patients with pCR were defined according to the tumor regression grade system.[18]

2.3 | Image acquisition and collagen feature extraction

MPI and collagen structural feature extraction were as follows: A ×20 objective lens was selected in this study to image the entire biopsy tissue and present the collagen structural features.[12,19] Then, multiphoton images were compared with the H&E images for histological evaluation. The extraction of collagen features was performed using MATLAB 2016b (MathWorks).[20] In total, 142 collagen features were extracted, including eight morphological features and 134 texture features (Table S1). More details about the imaging system and feature extraction are provided in the Supplementary Information.

2.4 | Collagen feature selection and collagen signature construction

Least absolute shrinkage and selection operator (Lasso) regression is characterized by variable selection and complexity regularization while fitting the generalized linear model. It can be used to select the most predictive markers from high-dimensional data and reduce the interaction between markers to avoid overfitting.[21] Therefore, Lasso regression was used to select collagen features and construct the collagen signature.

2.5 | Development and validation of the individualized prediction model

Univariable and multivariable logistic regression analyses were used to analyze the value of clinicopathological candidate predictors and collagen signatures in the primary cohort. Then, an individualized prediction model for pCR was developed based on the results of the multivariable analysis and presented as a visual nomogram.[22,23] The discrimination and calibration of the nomogram was measured by the ROC curve and calibration curve with the Hosmer–Lemeshow test. In addition, the variance inflation factor was calculated to evaluate the multicollinearity of the multivariate prediction model.

Internal validation was performed by the bootstrap method in the primary cohort to calculate the mean concordance index. The performance of the nomogram was validated in the independent validation cohort, and the ROC curves, calibration curves, and Hosmer–Lemeshow tests were assessed.

2.6 | Clinical utility of the prediction model

DCA and CIC were used to assess the clinical usefulness of the nomogram.[24] In addition, all patients were divided into two groups according to the Youden index in the primary cohort, namely, the high- and low-probability pCR groups, to assess the sensitivity,

specificity, accuracy, PPV and NPV of the prediction model in the primary cohort, validation cohort, and all patients, respectively.

2.7 | Incremental value of the collagen signature to traditional model

To estimate the incremental value of the collagen signature to the clinicopathological predictors, a clinicopathologic characteristic-based model (i.e., the traditional model) was developed without a collagen signature for comparison with the nomogram. Furthermore, the improvement of the nomogram based on the collagen signature was evaluated by the area under the ROC (AUC), NRI, and index IDI.

$$
\begin{aligned}
\text{Collagen signature} \ = -\ & 150.079256 + 152.701843 * \text{collagen straightness} \\
+\ & 112.973881 * \text{collagen crosslink density} + 9.718559 * \text{collagen orientation} \\
-\ & 1.017896 * \text{Gabor\_scale 4\_orientation 3\_mean}
\end{aligned}
$$

2.8 | Follow-up and association of the prediction model with prognosis

Patients achieved follow-up after radical surgery. The association between the nomogram-predicted high- and low-probability pCR and DFS and OS was analyzed.

2.9 | Statistical analysis

All statistical tests were performed using SPSS 24.0 and R statistical software (version 4.0.3). The chi-square test or Fisher's exact test was applied to compare categorical variables. Univariate and multivariate logistic regression analyses were used to identify the ORs of independent predictors and 95% confidence intervals (CIs). Survival curves are presented according to the Kaplan–Meier method and were compared by the log-rank test. A Cox proportional hazards model was used to determine the HR and 95% CI of variables for DFS and OS. Statistical tests were two-sided, and $p < 0.05$ was considered statistically significant.

3 | RESULTS

3.1 | Patient characteristics

According to the inclusion and exclusion criteria, 516 patients were included in this study (353 and 163 in the primary and validation cohorts, respectively) (Figure S1). The detailed baseline characteristics of the primary and validation cohorts is listed in Table 1. Univariate analysis revealed that differentiation status, pretreatment CEA level, pretreatment CA199 level, pretreatment T stage, and tumor dimension were significantly different between the pCR and non-pCR groups in the primary and validation cohort cohorts ($p < 0.05$).

The rate of pCR in the primary (21.5%, 76/353) and validation cohorts (22.7%, 37/163) was balanced ($p = 0.819$), and the baseline characteristics were similar (Table S2) between the two cohorts, which verified their use as primary and validation cohorts.

3.2 | Collagen feature selection and collagen signature construction

The flowchart of this research is presented in Figure 1. In total, 142 collagen features shrunk to four potential features by implementing Lasso regression in the primary cohort (Figure S2). These collagen features were presented in the collagen signature calculation formula:

Tumor budding was identified in the pretreatment biopsy in 70 of the 353 patients (19.8%). Comparison of the four collagen features with tumor budding showed statistically significant differences in collagen straightness, collagen crosslink density and collagen orientation (Table S3).

Representative H&E images, SHG/TPFF images, and binary images of the pCR and non-pCR patients are presented in Figure 2A–H. The distributions of the collagen signature of each patient in the two cohorts are presented in Figure 3. Obviously, patients with pCR had a lower collagen signature than non-pCR patients in the primary cohort (−0.830 [−2.059, −0.092] vs. 0.333 [−0.111, 0.771], $p < 0.001$), which was then verified in the validation cohort (−1.103 [−1.929, −0.199] vs. 0.321 [−0.135, 0.810], $p < 0.001$; Table 1). The collagen signature yielded an AUC of 0.842 (95% CI 0.788–0.895) in the primary cohort and 0.836 (95% CI 0.754–0.919) in the validation cohort (Figure 2I,J).

Furthermore, the collagen signature was significantly associated with pCR in the primary and validation cohorts when stratified analysis was performed (Tables S4,S5; Figures S3,S4).

3.3 | Development of the individualized prediction model

Differentiation status (OR: 2.529, 95% CI 1.216–5.259; $p = 0.013$), pretreatment CEA level (OR: 2.620, 95% CI 1.260–5.449; $p = 0.010$), pretreatment T stage (OR: 2.783, 95% CI 1.373–5.642; $p = 0.005$), tumor dimension (OR: 3.609, 95% CI 1.249–10.434, $p = 0.018$; OR: 5.347, 95% CI 1.371–20.849; $p = 0.016$), and collagen signature (OR: 0.269, 95% CI 0.181–0.400; $p < 0.001$) were identified as independent predictors for predicting pCR by multivariable analysis (Table 2). A prediction model that integrated these five predictors was constructed and presented as a nomogram (Figure 4A). Among these independent predictors, the

TABLE 1 Clinicopathological characteristics of the patients in the primary and validation cohorts

Characteristic	Primary cohort (n = 353)			Validation cohort (n = 163)		
	pCR (n = 76)	Non-pCR (n = 277)	p	pCR (n = 37)	Non-pCR (n = 126)	p
Age, years old			0.687			0.425
<60	51 (67.1)	179 (64.6)		19 (51.4)	74 (58.7)	
≥60	25 (32.9)	98 (35.4)		18 (48.6)	52 (41.3)	
Sex			0.576			0.418
Male	52 (68.4)	180 (65.0)		27 (73.0)	83 (65.9)	
Female	24 (31.6)	97 (35.0)		10 (27.0)	43 (34.1)	
BMI			0.140			0.670
<24	60 (78.9)	195 (70.4)		26 (70.3)	93 (73.8)	
≥24	16 (21.1)	82 (29.6)		11 (29.7)	33 (26.2)	
Differentiation status			<0.001			<0.001
Moderate or poor	40 (52.6)	226 (81.6)		21 (56.8)	106 (84.1)	
Well	36 (47.4)	51 (18.4)		16 (43.2)	20 (15.9)	
Pretreatment CEA level						0.003
Elevated	18 (23.7)	145 (52.3)	<0.001	11 (29.7)	74 (58.7)	
Normal	58 (76.3)	132 (47.7)		26 (70.3)	52 (41.3)	
Pretreatment CA199 level			0.027			0.109
Elevated	11 (14.5)	74 (26.7)		5 (13.5)	33 (26.2)	
Normal	65 (85.5)	203 (73.3)		32 (86.5)	93 (73.8)	
Distance from anal verge, cm			0.099			0.366
≤5	15 (19.7)	81 (29.2)		7 (18.9)	33 (26.2)	
>5	61 (80.3)	196 (70.8)		30 (81.1)	93 (73.8)	
Pretreatment T stage			0.006			0.011
T4	18 (23.7)	113 (40.8)		13 (35.1)	74 (58.7)	
T3	58 (76.3)	164 (59.2)		24 (64.9)	52 (41.3)	
Pretreatment N stage			0.113			0.449[a]
N+	58 (76.3)	233 (84.1)		33 (89.2)	105 (83.3)	
N−	18 (23.7)	44 (15.9)		4 (10.8)	21 (16.7)	
Tumor dimension, cm			<0.001			0.008
>5	6 (7.9)	79 (28.5)		6 (16.2)	43 (34.1)	
>3, ≤5	57 (75.0)	196 (64.3)		25 (67.6)	78 (61.9)	
≤3	13 (17.1)	20 (7.2)		6 (16.2)	5 (4.0)	
Collagen signature, median (IQR)	−0.830 (−2.059, 0.092)	0.333 (−0.111, 0.771)	<0.001	−1.103 (−1.929, 0.199)	0.321 (−0.135, 0.810)	<0.001

Note:: Values in parentheses are percentages unless indicated otherwise.

The *p* value is derived from the univariable association analyses between each of the clinicopathological characteristics and pCR.

Abbreviations: BMI, body mass index; CA199, carbohydrate antigen 199; CEA, carcinoembryonic antigen; IQR, interquartile range; pCR, pathological complete response.

[a]Fisher's exact test.

collagen signature had the most discriminative ability for predicting pCR (Figure 2I,J; Table S6). The variance inflation factor of the five predictors was less than five, demonstrating no multicollinearity among all predictors (Figure S5).

We further investigated the relationship between these four clinicopathological predictors and the four collagen features (Tables S7–S10). The results showed that four collagen features in the tumor microenvironment were associated with clinicopathological characteristics: (1) moderate or poor differentiation was related to high collagen orientation; (2) elevated pretreatment CEA level was associated with high collagen orientation; (3) patients with cT4 stage have higher collagen orientation and lower Gabor feature compared with cT3 stage; and (4) tumor dimension was significantly correlated with collagen orientation and Gabor feature.
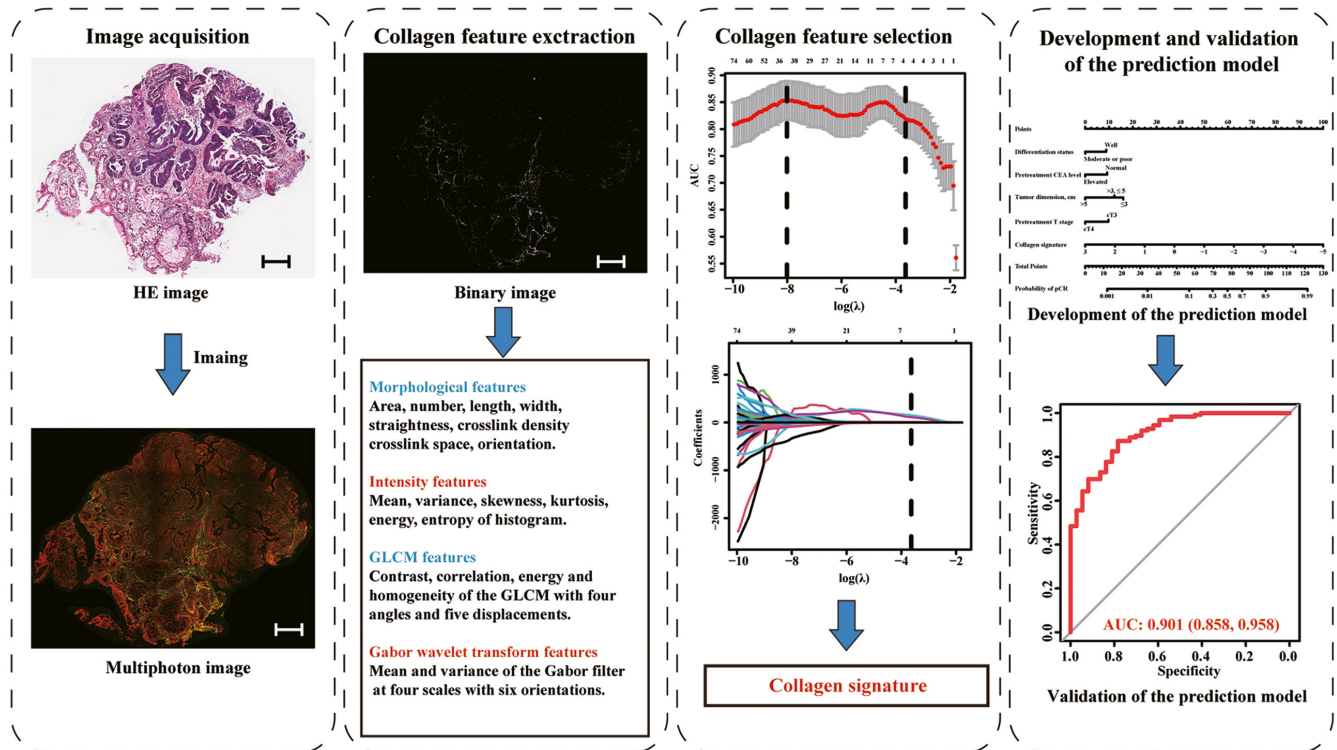
FIGURE 1 Flowchart of this study. This study included image acquisition, collagen feature extraction, collagen feature selection, collagen signature construction, and development and validation of the prediction model. The images in the left panel are representative H&E images and the corresponding multiphoton images, including TPEF and SHG, of the pretreatment biopsy (collagen is presented in green). Then, the SHG signal image is converted into the binary image (collagen is presented as white) for collagen feature extraction. Next, Lasso logistic regression was applied to select predictive collagen features to construct a collagen signature. A prediction model integrating the collagen signature and clinicopathological predictors was developed and assessed in the validation cohort. Scale bars: 200 μm. AUC, area under the curve; CEA, carcinoembryonic antigen; GLCM, gray-level co-occurrence matrix; pCR, pathological complete response; SHG, second harmonic generation; TPEF two-photon excitation fluorescence

3.4 | Evaluation and validation of the performance of the prediction model

The AUC of the nomogram was 0.891 (95% CI 0.847–0.935) in the primary cohort (Figure 4B), and the calibration curve demonstrated good agreement between the predicted pCR probability and actual pCR probability in the primary cohort (Figure 4C). The internal validation results showed a mean concordance index of 0.893 by the bootstrap method.

Good performance (Figure 4D) and favorable calibration (Figure 4E) of the nomogram were also verified in the validation cohort, with an AUC of 0.908 (95% CI 0.858–0.958). The Hosmer–Lemeshow test showed nonsignificant statistics in the primary cohort ($p = 0.314$) and the validation cohort ($p = 0.670$), proving that there was goodness of fit of the prediction model.

3.5 | Clinical utility of the prediction model

DCA indicated that using the nomogram to predict pCR showed a greater advantage than either the "treat-all scheme" or "treat-none scheme" in the primary cohort, validation cohort, and all patients

(Figure 5A). Based on these DCAs, CICs were performed to evaluate the clinical impact of the nomogram to help us more intuitively recognize its significant value by building a simulated model comprised of 1000 LARC cases to more accurately identify patients with potential pCR. The results showed the great predictive ability of the nomogram when the probability threshold of nearly 0.4 was optimal to identify patients who would achieve pCR from nCRT (Figure 5B).

In addition, the maximum value of the Youden index was 0.251, which was the cut-off value in the primary cohort. Then, the patients were separated into a high probability pCR group and a low-probability pCR group. The nomogram also had satisfactory sensitivity, specificity, accuracy, PPV, and NPV (Table 3).

3.6 | Incremental value of the collagen signature to traditional model

The collagen signature was excluded, and a traditional model based on pretreatment CEA level, pretreatment CA199 level, differentiation status, pretreatment T stage, and tumor dimension (Table S11) was developed. The AUCs of the traditional model were 0.804 (95% CI 0.704–0.860) in the primary cohort, 0.789 (95% CI 0.706–0.872)
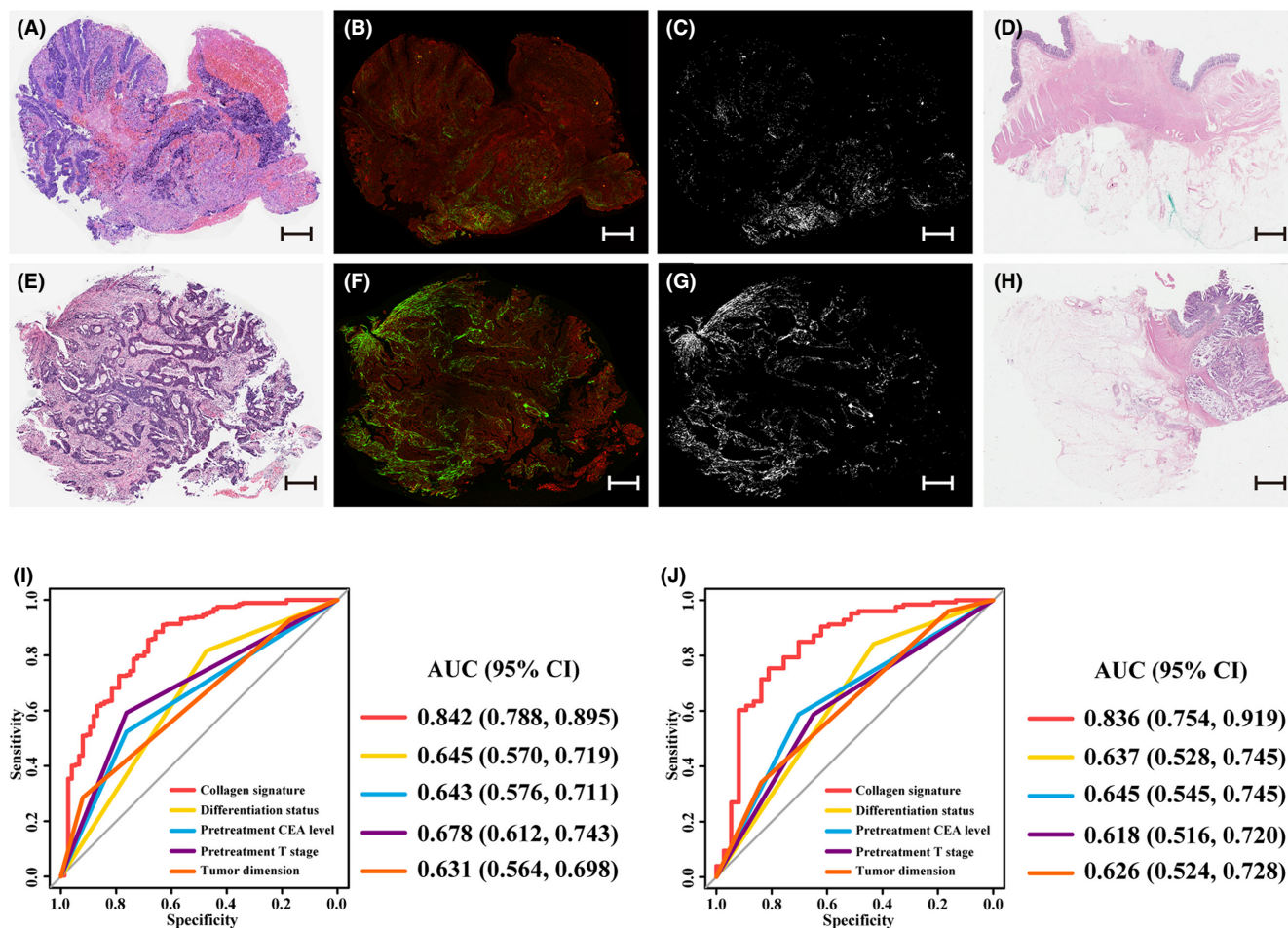
FIGURE 2 Representative images of tissues from patients with pCR and non-pCR and ROC curves of the predictors for predicting pCR. (A–D) From left to right are the representative H&E images, corresponding multiphoton image (collagen is presented in green), and binary image (collagen is presented in white) of a pretreatment biopsy and H&E image of a postoperative resection specimen for patients with pCR. (E–H) From left to right are the representative H&E image, corresponding multiphoton image (collagen is presented in green), and binary image (collagen is presented in white) of a pretreatment biopsy and H&E image of a postoperative resection specimen for patients with non-pCR. The ROC curves of the collagen signature and clinicopathological predictors in the primary cohort (I) and the validation cohort (J). Scale bars: (A–C) 200 μm and (d) 2 mm; (E–G) 200 μm; and (H) 2 mm. AUC, area under the curve; CA199, carbohydrate antigen199; CEA, carcinoembryonic antigen; pCR, pathological complete response; ROC, receiver operating characteristic curve

in the validation cohort, and 0.799 (95% CI 0.752–0.846) in all patients. The nomogram demonstrated better discrimination ability for predicting pCR than the traditional model (Table 4; Figure 5C). Moreover, all the NRI and IDI values were >0, with p-values <0.05 between the nomogram and traditional model, indicating that the nomograms performed better than the traditional model (Table 4). DCA also showed that the nomogram had a higher net benefit than the traditional model for predicting the probability of pCR (Figure 5A). In addition, the nomogram had higher sensitivity, specificity, accuracy, PPV, and NPV than the traditional model (Table 3).

3.7 | Follow-up and association of the prediction model with prognosis

The median (IQR) DFS and OS were 44.5 months (28–57 months) and 48 months (36–58 months), respectively. Among patients with

a high probability of pCR, DFS was significantly better than that among patients with a low probability of pCR (3-year DFS: high probability of pCR, 91.2%; low probability of pCR, 70.6%; log-rank $p < 0.001$; Figure 6A). Furthermore, the OS of patients with a high probability of pCR was also better than that of patients with a low probability of pCR (3-year OS: high probability of pCR, 94.1%; low probability of pCR, 81.3%; log-rank $p < 0.001$; Figure 6B). The collagen signature and other predictors with the corresponding survival status are shown in Figure 7.

Cox proportional hazard regression found that differentiation status (HR: 0.442, 95% CI 0.232–0.842; $p = 0.013$), pretreatment CEA level (HR: 1.561, 95% CI 1.061–2.299; $p = 0.024$) and nomogram-predicted probability of pCR (HR: 2.475, 95% CI 1.308–4.682; $p = 0.005$) were independent prognostic factors for DFS. Similarly, differentiation status (HR: 0.511, 95% CI 0.238–1.097; $p = 0.085$) and nomogram-predicted probability of pCR (HR: 2.792, 95% CI 1.351–5.770; $p = 0.006$) were independent prognostic factors
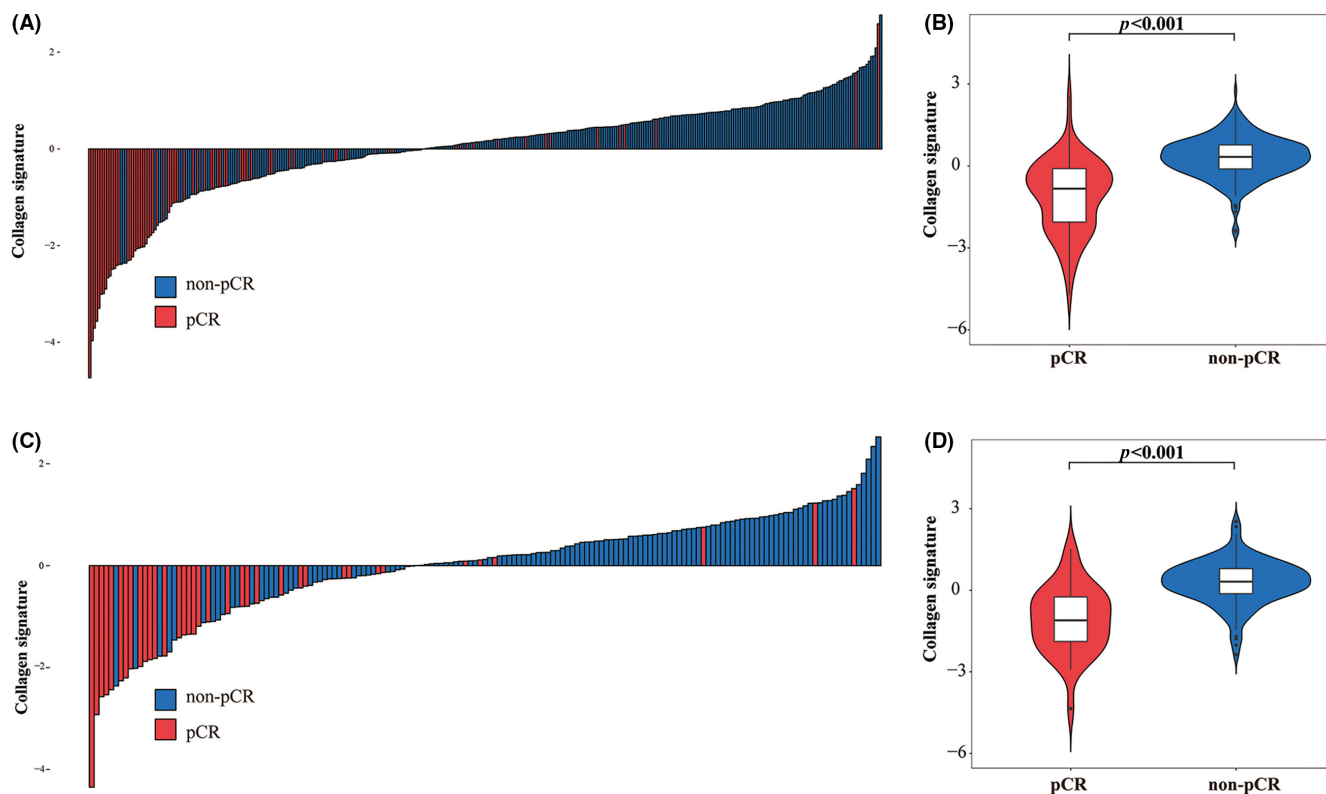
FIGURE 3 Distribution of the collagen signature in the primary cohort and the validation cohort. (A, B) Collagen score for each patient in the validation cohort and comparison of the collagen signature between patients with pCR and non-pCR in the primary cohort. (C, D) Collagen score for each patient in the primary cohort and comparison of the collagen signature between patients with pCR and non-pCR in the primary cohort. Red represents pCR, and blue represents non-pCR. pCR, pathological complete response

for OS (Table 5). This result showed that the nomogram-predicted probability of pCR was significantly associated with prognosis after adjusting for other variables.

4 | DISCUSSION

In this study, we developed and validated a prediction model based on the collagen signature and presented it as an easy-to-use nomogram. The nomogram with satisfactory performance was intended to be used by surgeons to predict the personalized probability of pCR and provide an effective tool for clinical decision-making.

Collagen is the main component of the ECM; it provides structural and mechanical support for cells and tissues and regulates a variety of cell functions.[25] Growing evidence has proven that changes in the collagen structure in the tumor microenvironment could importantly influence the growth, invasion, metastasis, and survival of tumor cells and even affect therapeutic sensitivity.[26-28] Therefore, collagen with great potential clinical application value is currently one of the hotspots of individualized medical research.[29-31] However, the relationship between pCR and collagen structure in the tumor microenvironment of pretreatment biopsy is unclear. With the development of interdisciplinary approaches, MPI can accurately and selectively be used to visualize collagen in the tumor microenvironment in a label-free manner.[32-34] In addition, it is feasible to automatically

extract high-throughput collagen feature information from multiphoton images for conducting subsequent data analysis to provide decision support.[14,15] Based on the above factors, pretreatment biopsy was imaged by MPI, and 142 collagen features were extracted in this study for subsequent analysis.

In recent studies, multimarker analyses that combine singular markers into marker panels have been accepted and can increase the prediction performance.[35] Lasso regression is a useful algorithm to select the most predictive value of parameters from high-dimensional data while avoiding overfitting.[36,37] In this study, we shrank the regression coefficients by Lasso regression to build the collagen signature. As a result, 142 candidate collagen features were reduced to four potential predictors: collagen straightness, collagen crosslink density, collagen orientation, and Gabor_scale4_orientation3_mean. Then, the collagen signature was constructed. The collagen signature demonstrated good discrimination in the primary cohort (AUC = 0.842), which was then confirmed in the validation cohort (AUC = 0.836). Therefore, the collagen signature based on four collagen features of pretreatment biopsy was significantly associated with pCR in LARC patients. The morphologically predictive collagen features were easily identified. Previous studies have proven that straightness, crosslink density, and orientation of collagen are associated with treatment resistance and tumor proliferation,[38-41] which is consistent with our results; that is, high values of straightness, crosslink density, and orientation have a higher

TABLE 2 Univariate and multivariate analyses of the predictors of pCR in the primary cohort

Variables	Univariate analysis		Multivariate analysis	
	OR (95% CI)	p	OR (95% CI)	p
Age, years old		0.687		
<60	Reference			
≥60	0.895 (0.523, 1.534)			
Sex		0.576		
Male	Reference			
Female	0.856 (0.498, 1.474)			
BMI		0.143		
<24	Reference			
≥24	0.634 (0.345, 1.166)			
Differentiation status		<0.001		0.013
Moderate or poor	Reference		Reference	
Well	3.988 (2.317, 6.866)		2.529 (1.216, 5.259)	
Pretreatment CEA level		<0.001		
Elevated	Reference		Reference	0.010
Normal	3.540 (0.984, 6.315)		2.620 (1.260, 5.449)	
Pretreatment CA199 level		0.030		NA
Elevated	Reference		NA	
Normal	2.154 (1.078, 4.304)		NA	
Distance from anal verge, cm		0.102		
≤5	Reference			
>5	1.681 (0.903, 3.128)			
Pretreatment T stage		<0.001		0.005
T4	Reference		Reference	
T3	4.676 (2.617, 8.357)		2.783 (1.373, 5.642)	
Pretreatment N stage		0.116		
N+	Reference			
N−	1.643 (0.885, 3.053)			
Tumor dimension, cm				
>5	Reference		Reference	
>3, ≤5	4.216 (1.745, 10.185)	0.001	3.609 (1.249, 10.434)	0.018
≤3	8.558 (2.893, 25.319)	<0.001	5.347 (1.371, 20.849)	0.016
Collagen signature	0.226 (0.156, 0.328)	<0.001	0.269 (0.181, 0.400)	<0.001

Abbreviations: BMI, body mass index; CA199, carbohydrate antigen199; CEA, carcinoembryonic antigen; CI, confidence interval; NA, not available; OR, odds ratio.

collagen signature with a low probability of pCR after nCRT. In addition, another feature used to construct the collagen signature was a Gabor wavelet transform feature. The Gabor wavelet transform is a multiscale image analysis method that divides the image data into different frequency components.[42] The texture features extracted from the wavelet decomposed image can further present the spatial heterogeneity of collagen at multiple scales.[43] Recently, Chen et al.[10] developed a prediction model based on four collagen features to predict the peritoneal metastasis of gastric cancer, among which three were Gabor wavelet transform features. This study confirmed that the Gabor wavelet transform feature is closely associated with

the biological behavior of tumors and is a critical imaging biomarker for predicting therapeutic efficacy and prognosis.

The presence of tumor budding is associated with a poor treatment response to nCRT and showed more aggressive tumor behavior in pretreatment biopsy samples. In this work, we found significant differences, including collagen straightness, collagen crosslink density, and collagen orientation, between the patients with tumor budding and without tumor budding. Several studies have also reported that collagen straightness, collagen crosslink density, and collagen orientation are associated with tumor aggressiveness[38-40]; straighter collagen in the tumor microenvironment is a highway for tumor cell
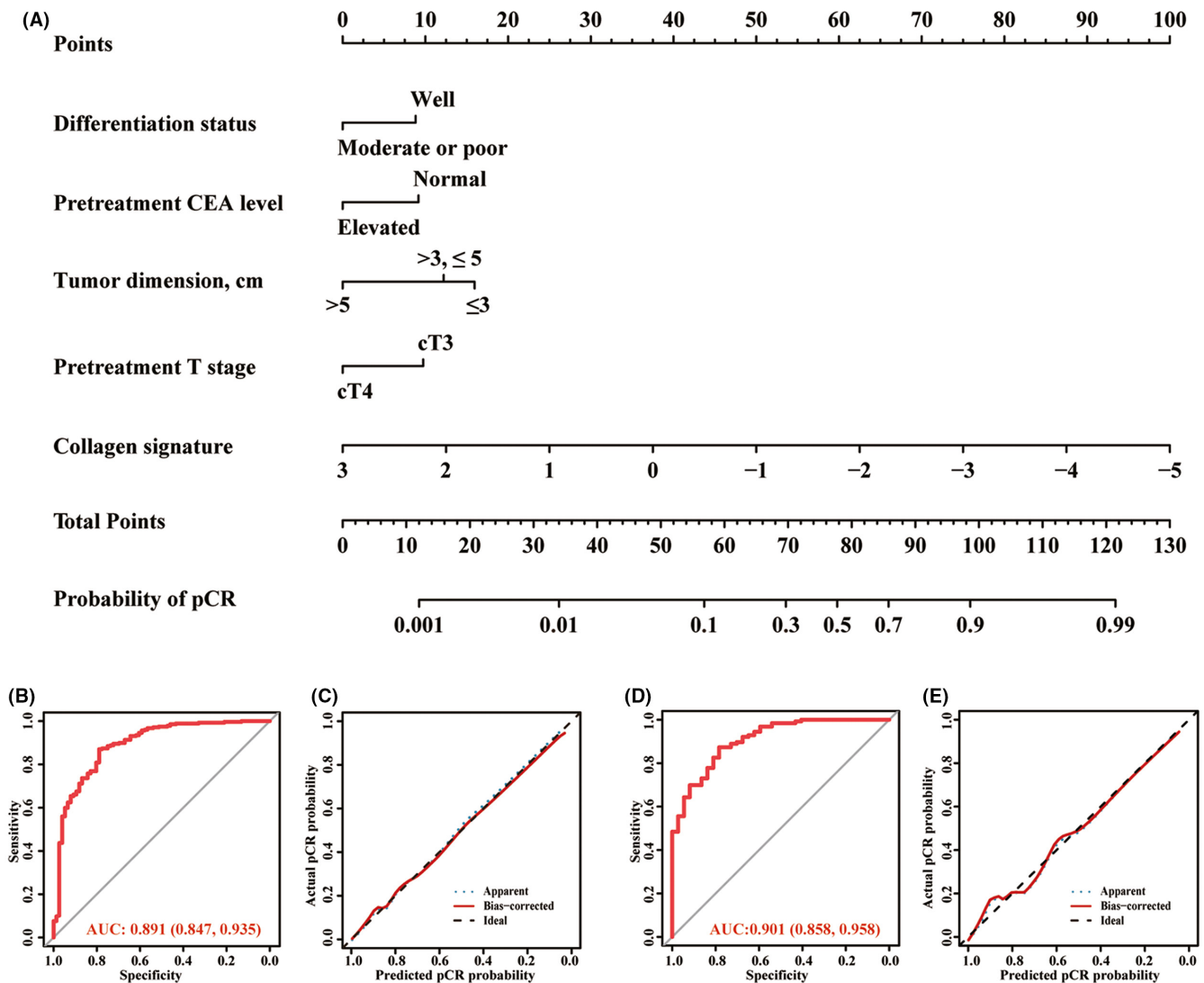
(A)

FIGURE 4 Development, validation, and evaluation of the performance of the nomogram in the primary cohort and the validation cohort. (A) The nomogram is developed in the primary cohort, with the collagen signature, differentiation status, pretreatment CEA level, pretreatment T stage, and tumor dimension incorporated. (B, C) The ROC curve and the calibration curve of the nomogram in the primary cohort. (D, E) The ROC curve and the calibration curve of the nomogram in the validation cohort. In the calibration curve, the y-axis represents the actual pCR probability, and the x-axis represents the predicted pCR probability. The diagonal black dotted line represents a perfect prediction in an ideal model. The solid red line is a representation of the collagen nomogram; better prediction is indicated when the solid red line has a closer fit to the diagonal black dotted line. AUC, area under the curve; CEA, carcinoembryonic antigen; pCR, pathological complete response; ROC, receiver operating characteristic curve

migration.[44] Therefore, these results suggested that tumor cells are prone to migrate and develop tumor budding in the tumor microenvironment with high collagen straightness, collagen crosslink density, and collagen orientation.

The epithelial–mesenchymal transition (EMT), the process during which epithelial cells lose adhesion with neighboring cells and are converted to migratory and invasive cells, is closely tied to cancer progression.[45] Collagen in the ECM is critical for EMT.[46] Four collagen features may reflect the tumor microenvironment, which is related to the promotion of EMT. Patients with high collagen straightness, collagen orientation, collagen crosslink density, and low Gabor feature may represent increased matrix stiffness.[44] Increased matrix stiffness could lead to improved interstitial pressure, tumor and stromal

cell deformation, and initiation of EMT.[46] Moreover, increased matrix stiffness could also drive EMT through a TWIST1-G3BP2 mechano-transduction pathway.[47] The increased collagen crosslink density can promote EMT by weakening cell–cell adhesions.[48] Furthermore, the increased crosslink density indicated collagen deposition. Deposition of collagen can promote EMT through various biological pathways, such as enhancement of Snail stability through discoid domain receptor 2,[49] disruption of E-cadherin,[50] and SMADS.[51]

Currently, the accurate prediction of pCR using traditional clinicopathological characteristics remains challenging in clinical settings. To show the incremental value of the collagen signature compared with the traditional model, we excluded the collagen signature and built a traditional model. The results showed that the
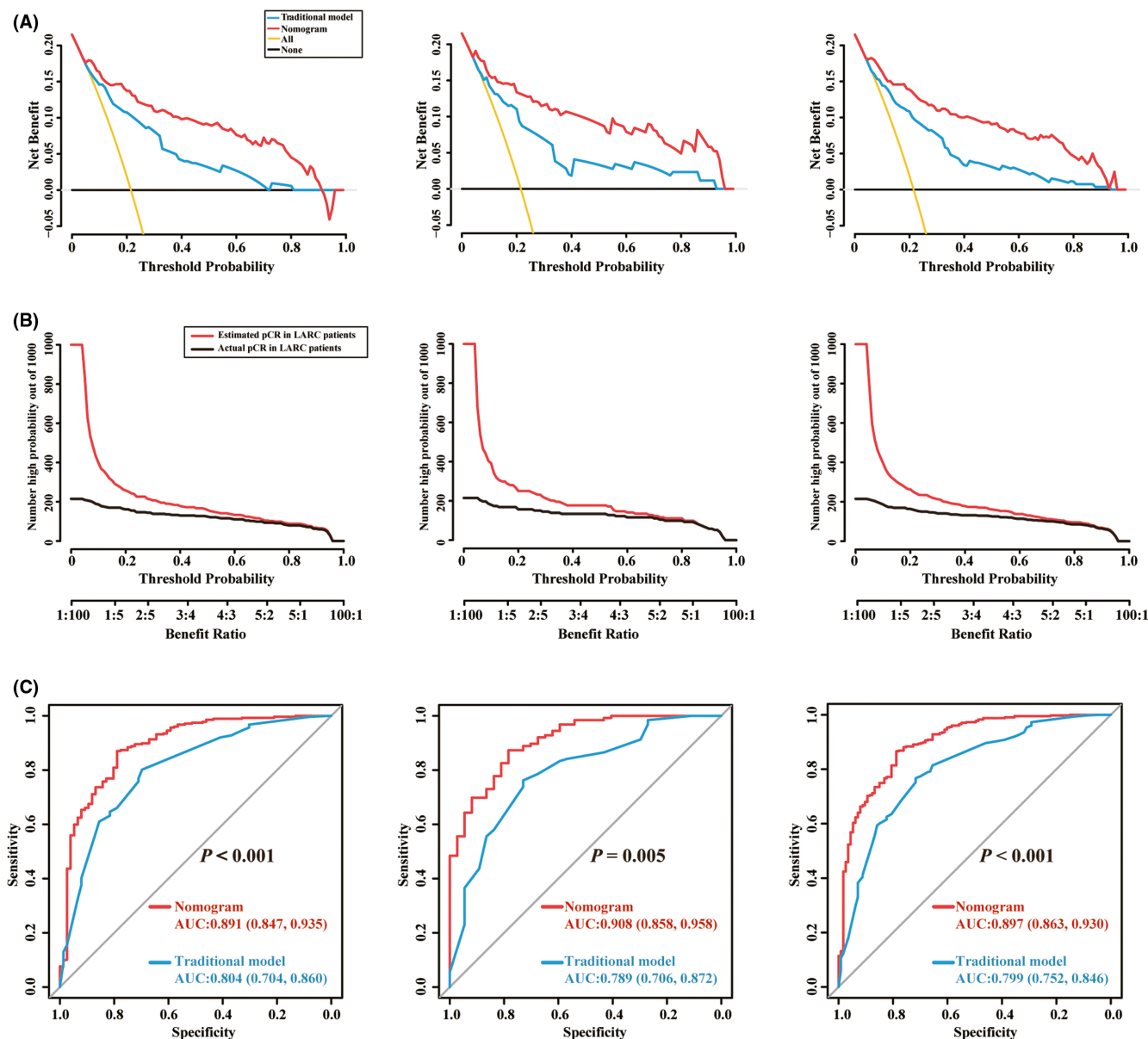
FIGURE 5 Clinical utility of the nomogram. From left to right are the primary cohort, validation cohort, and all patients. (A) Decision curve analysis for the nomogram. The *y*-axis represents the net benefit, the *x*-axis represents the different threshold probabilities, the red line represents the collagen nomogram, the cyan line represents the traditional model, the yellow line represents the "treat-all scheme," and the black line represents the "treat-none scheme." The decision curve revealed that using the nomogram to predict pCR could add more benefit than the traditional model, the "treat-all scheme" and the "treat-none scheme." (B) Clinical impact curves for the nomogram. Of 1000 patients, the red line shows the total number of LARC patients who would be deemed pCR for each threshold probability. The black line shows how many of those would be true positives (cases). The closer the curves, the higher the probability that the nomogram would identify pCR patients from a total estimated number of pCR in LARC patients. The threshold value represents the value after which the rate of misdiagnosis would be lowest, thereby providing an optimal benefit ratio for the patient. (C) ROCs for the nomogram and the traditional model. The red line represents the nomogram; the cyan line represents the traditional model. AUC, area under the curve; LARC, locally advanced rectal cancer; pCR, pathological complete response; ROC, area under the receiver operator characteristic

nomogram based on the collagen signature had higher AUC, sensitivity, specificity, accuracy, PPV, and NPV values than the traditional model. Improvements in NRI and IDI values indicated that the nomogram was superior to the traditional model. Moreover, DCA curves indicated that the nomogram had better applicability than the traditional model. Based on the above results, the incremental value of the collagen signature could be definitely identified.

Some studies have used radiomics to predict pCR in rectal cancer. Liu et al.[1] used radiomic analysis to evaluate pCR with a very high AUC of 0.975, but this model was developed before and after treatment imaging; therefore, it cannot predict pCR before treatment. Nie et al.[52] used pretreatment multiparametric MRI images to predict pCR and obtained an AUC of 0.84. The radiomics method depends on domain expertise to manually mark handcrafted features,

TABLE 3 The performance of the nomogram and the traditional model in predicting pCR in the primary cohort, validation cohort, and all patients

Variables	Primary cohort		Validation cohort		All patients	
	Nomogram	Traditional model	Nomogram	Traditional model	Nomogram	Traditional model
Cut-off	0.251	0.321	0.251	0.321	0.251	0.321
Sensitivity, %	78.9 (68.5–86.6)	67.1 (55.9–76.6)	62.8 (78.4–88.6)	56.8 (40.9–71.3)	78.8 (70.3–85.3)	63.7 (54.5–72.0)
Specificity, %	87.0 (82.5–80.5)	81.2 (76.2–85.4)	85.7 (78.5–90.8)	84.1 (76.8–89.5)	86.6 (82.9–89.6)	82.1 (78.1–85.6)
Accuracy, %	85.3 (81.2–88.6)	78.2 (73.6–85.4)	84.0 (77.7–88.9)	77.9 (70.9–83.6)	84.9 (81.5–87.7)	78.1 (74.3–81.5)
PPV, %	62.5 (52.5–71.5)	49.5 (40.1–59.0)	61.7 (47.4–74.2)	51.2 (36.5–65.7)	62.2 (54.1–69.8)	50.0 (41.9–58.1)
NPV, %	93.8 (90.1–96.1)	90.0 (85.7–93.1)	93.1 (87.0–96.5)	86.9 (79.8–91.8)	93.6 (80.6–95.6)	89.0 (85.4–91.8)

Note: Values are percentages unless indicated otherwise.

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

TABLE 4 Performance comparison between the nomogram and the traditional model

Variables	AUC (95% CI)	p	NRI (95% CI)	p	IDI (95% CI)	p
Primary cohort		<0.001		0.001		<0.001
Nomogram	0.891 (0.847, 0.935)		0.180 (0.074, 0.286)		0.209 (0.146, 0.272)	
Traditional model	0.804 (0.704, 0.860)		Reference		Reference	
Validation cohort		0.005		0.033		<0.001
Nomogram	0.901 (0.858, 0.958)		0.216(0.017, 0.416)		0.247 (0.154, 0.339)	
Traditional model	0.789 (0.706, 0.872)		Reference		Reference	
All patients		<0.001		0.003		<0.001
Nomogram	0.897 (0.863, 0.930)		0.158 (0.055, 0.260)		0.222 (0.171, 0.274)	
Traditional model	0.799 (0.752, 0.846)		Reference		Reference	

Abbreviations: AUC, area under the curve; CEA, carcinoembryonic antigen; CI, confidence interval; IDI, integrated discrimination improvement; NRI, net reclassification improvement.
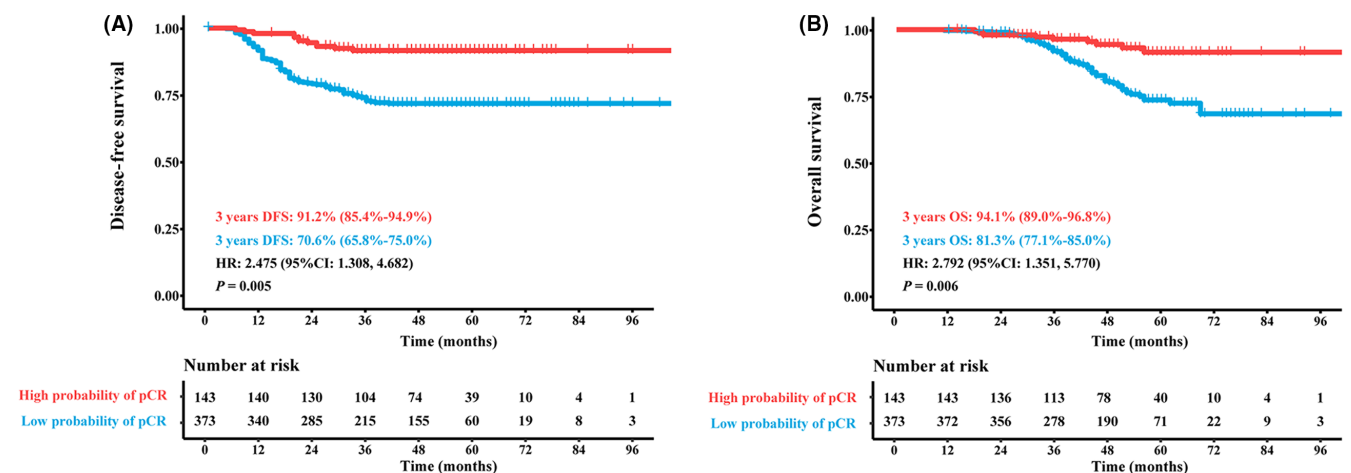

FIGURE 6 Kaplan–Meier analysis of disease-free survival and OS according to the nomogram-predicted subgroups of all patients. (A) Disease-free survival of all patients in the high- and low-probability pCR subgroups. (B) OS of all patients in the high- and low-probability pCR subgroups. pCR, pathological complete response

and radiomics also needs accurate tumor segmentation.[53] The collagen signature we proposed is based on biopsy specimens before treatment, so the error caused by the manual marking process is avoided. Furthermore, the collagen signature provided additional prognostic information and helped researchers to understand the interactions between tumor cells and their structural microenvironment; here, the collagen signature was clinically relevant and worth investigating.
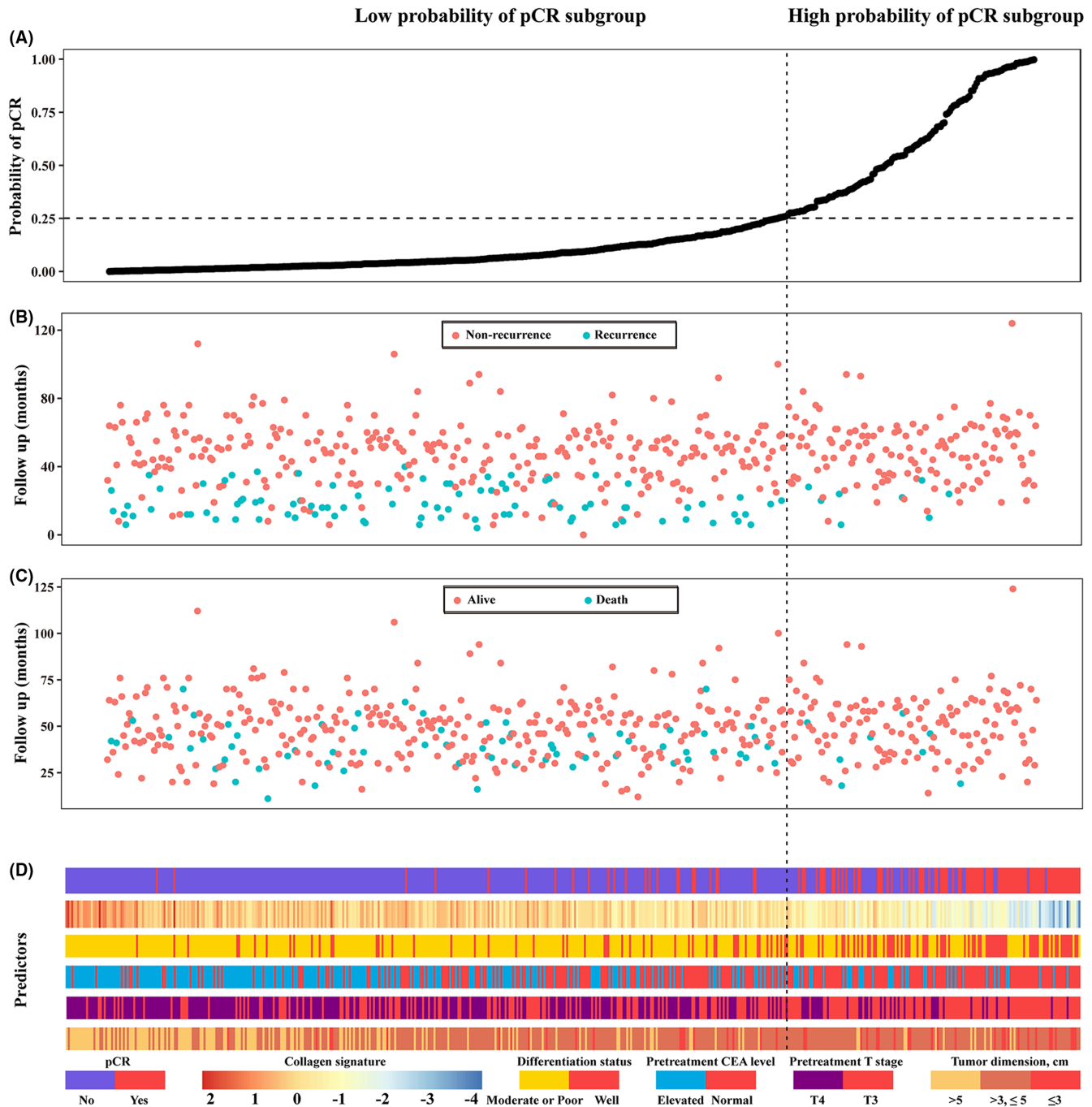
FIGURE 7 Distribution of the nomogram-predicted subgroups with the corresponding survival status in all patients. (A) Nomogram-predicted probability of pCR distribution. (B) Disease-free survival status of all patients. (C) OS status of all patients. (D) Distribution of the collagen signature and clinicopathological predictors with the corresponding survival status. The black dotted line represents the cut-off dividing the patients into high- and low-probability pCR groups. CEA, carcinoembryonic antigen; pCR, pathological complete response

We suggest that the simpler procedure is used in routine practice as follows. First, biopsy is performed in the diagnosis of LARC before nCRT. Second, the H&E or unstained section of biopsy tissue is subjected to multiphoton imaging after a routine pathology procedure, which can be completed in only 2–3 min. Third, the collagen feature was automatically extracted from the multiphoton image by MATLAB software within 1 min; then, the collagen signature was calculated. Finally, the probability of pCR was obtained through the

nomogram before nCRT. Therefore, the prediction model could potentially be used to predict pCR in rectal cancer patients who may safely adopt the "wait and see" policy after nCRT and help surgeons communicate with patients for decision-making.

There are some limitations in this study. First, this study was a retrospective cohort, and selection bias cannot be avoided. Second, some studies have demonstrated that desmoplastic reaction (DR) classification is associated with the prognosis of

TABLE 5 Cox regression analysis of the preoperative predictors for survival in all patients

Variables	Univariate analysis HR (95% CI)	p	Multivariate analysis HR (95% CI)	p
Disease-free survival				
Age (years old) (≥60 vs. <60)	1.215 (0.827, 1.785)	0.320		
Sex (male vs. female)	1.057 (0.723, 1.456)	0.774		
BMI (≥24 vs. <24)	1.140 (0.767, 1.693)	0.517		
Differentiation status (well vs. moderate or poor)	0.307 (0.165, 0.571)	<0.001	0.442 (0.232, 0.842)	0.013
Pretreatment CEA (elevated vs. normal)	1.978 (1.358, 2.881)	<0.001	1.561 (1.061, 2.299)	0.024
Pretreatment CA199 (elevated vs. normal)	1.319 (0.885, 1.966)	0.174		
Distance from anal verge, cm (≤5 vs. >5)	1.140 (0.762, 1.705)	0.525		
Pretreatment T stage (T4 vs. T3)	1.426 (0.985, 2.063)	0.060	NA	NA
Pretreatment N stage (N+ vs. N−)	1.163 (0.732, 1.850)	0.522		
Tumor dimension, cm				
>5	Reference			
>3, ≤5	0.816 (0.546, 1.218)	0.320		
≤3	0.571 (0.253, 1.285)	0.175		
Nomogram-predicted (low vs. high probability)	3.990 (2.074, 6.851)	<0.001	2.475 (1.308, 4.682)	0.005
Overall survival				
Age (years old) (≥60 vs. <60)	1.145 (0.734, 1.786)	0.552		
Sex (male vs. female)	1.143 (0.720, 1.814)	0.571		
BMI (≥24 vs. <24)	1.261 (0.770, 2.065)	0.356		
Differentiation status (well vs. moderate or poor)	0.339 (0.163, 0.701)	0.004	0.511 (0.238, 1.097)	0.085
Pretreatment CEA (elevated vs. normal)	1.250 (0.815, 1.915)	0.307		
Pretreatment CA199 (elevated vs. normal)	1.137 (0.694, 1.862)	0.611		
Distance from anal verge, cm (≤5 vs. >5)	1.364 (0.868, 2.145)	0.179		
Pretreatment T stage (T4 vs. T3)	1.897 (1.211, 2.970)	0.005	NA	NA
Pretreatment N stage (N+ vs. N−)	1.135 (0.659, 1.955)	0.648		
Tumor dimension, cm				
>5	Reference			
>3, ≤5	0.984 (0.601, 1.611)	0.948		
≤3	0.936 (0.400, 2.192)	0.879		
Nomogram-predicted (low vs. high probability)	3.525 (1.766, 7.036)	<0.001	2.792 (1.351, 5.770)	0.006

Abbreviations: BMI, body mass index; CA199, carbohydrate antigen199; CEA, carcinoembryonic antigen; CI, confidence interval; HR, hazard ratio; NA, not available.

colorectal cancer.[54,55] However, we cannot clarify the relationship between these four collagen features and the DR category due to the endoscopic forceps being unable to obtain tumor biopsy tissue from the intrinsic muscle layer. Third, the purpose of this study was to develop a pretreatment collagen signature based on collagen features from the pretreatment biopsy to predict pCR. Therefore, we used pretreatment biopsy tissue rather than

posttreatment resected samples. We did not evaluate the discrepancy between the biopsy specimen and the resected sample in this study. Of course, the resected specimen underwent radiotherapy, which may cause excessive collagen deposition and structural disorganization by myosin IIA expression and oxidative stress.[56,57] In addition, the regressed tumor tissue is replaced by interstitial fibrosis.[58] In short, the collagen structure of the resected specimen

may be deposited and disorganized by radiotherapy compared with pretreatment biopsy tissue.

In conclusion, we found that the collagen signature in the tumor microenvironment of pretreatment biopsy samples was significantly associated with pCR. We developed and validated a nomogram based on the collagen signature for accurately individualized prediction of pCR in patients with LARC before nCRT.

AUTHOR CONTRIBUTIONS

WJ, SW, YG, SZ, and JY conceived and designed the study. WJ, SW, JW, JZ, XD, ZL, WX, YG, and JY acquired the data. WX, YG, and JY did the quality control of data. GW, SX, and SZ implemented the multiphoton imaging and collagen features extraction. YG and JY verified the data. All authors had access to, analyzed, and interpreted the data. WJ, SW, JW, and JZ did the statistical analyses. WJ, SW, JW, JZ, XD, and ZL developed, and validated the prediction model. WJ, SW, ZL, and JZ prepared the first draft of the manuscript. WJ, SW, JW, JZ, YG, SZ, and JY revised the manuscript. All authors contributed to manuscript preparation.

DISCLOSURE

The authors have no conflict of interest.

ETHICAL APPROVAL

Approval of the research protocol by an Institutional Reviewer Board: NFEC-2021-440; Registry and the Registration No. of the study/trial: N/A; Animal Studies: N/A.

ORCID

Yuanhong Gao https://orcid.org/0000-0002-6429-5376

Jun Yan https://orcid.org/0000-0003-1373-4380

REFERENCES

1. Liu Z, Zhang X-Y, Shi Y-J, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. 2017;23:7253-7262.
2. Brenner H, Kloor M, Pox C. Colorectal cancer. *Lancet*. 2014;383:1490-1502.
3. Zhang F, Yao SU, Li Z, et al. Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features. *Clin Transl Med*. 2020;10:e110.
4. Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg*. 2004;240:711-718; discussion 717–718.
5. van der Valk M, Hilling D, Bastiaannet E, et al. Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. *Lancet*. 2018;391:2537-2545.
6. Smith JJ, Strombom P, Chow OS, et al. Assessment of a watch-and-wait strategy for rectal cancer in patients with a complete response after neoadjuvant therapy. *JAMA Oncol*. 2019;5:e185896.
7. Bissell M, Hines W. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat Med*. 2011;17:320-329.
8. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nat Rev Cancer*. 2009;9:239-252.
9. Penet M-F, Kakkad S, Pathak AP, et al. Structure and function of a prostate cancer dissemination-permissive extracellular matrix. *Clin Cancer Res*. 2017;23:2245-2254.
10. Chen D, Liu Z, Liu W, et al. Predicting postoperative peritoneal metastasis in gastric cancer with serosal invasion using a collagen nomogram. *Nat Commun*. 2021;12:179.
11. Desa DE, Bhanote M, Hill RL, et al. Second-harmonic generation directionality is associated with neoadjuvant chemotherapy response in breast cancer core needle biopsies. *J Biomed Opt*. 2019;24:1-9.
12. Jiang W, Li M, Tan J, et al. A nomogram based on a collagen feature support vector machine for predicting the treatment response to neoadjuvant chemoradiotherapy in rectal cancer patients. *Ann Surg Oncol*. 2021;28:6408-6421.
13. Yan J, Zheng X, Liu Z, et al. Multiphoton imaging provides a superior optical biopsy to that of confocal laser endomicroscopy imaging for colorectal lesions. *Endoscopy*. 2019;51:174-178.
14. Xu S, Wang Y, Tai DCS, et al. qFibrosis: a fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients. *J Hepatol*. 2014;61:260-269.
15. Xu S, Kang CH, Gou X, et al. Quantification of liver fibrosis via second harmonic imaging of the Glisson's capsule from liver surface. *J Biophotonics*. 2016;9:351-363.
16. Rogers AC, Gibbons D, Hanly AM, et al. Prognostic significance of tumor budding in rectal cancer biopsies before neoadjuvant therapy. *Modern Pathol*. 2014;27:156-162.
17. Breugom AJ, van Gijn W, Muller EW, et al. Adjuvant chemotherapy for rectal cancer patients treated with preoperative (chemo) radiotherapy and total mesorectal excision: a Dutch Colorectal Cancer Group (DCCG) randomized phase III trial. *Ann Oncol*. 2015;26:696-701.
18. Karagkounis G, Thai L, Mace AG, et al. Prognostic implications of pathological response to neoadjuvant chemoradiation in pathologic stage III rectal cancer. *Ann Surg*. 2019;269:1117-1123.
19. Wang G, Sun Y, Chen Y, et al. Rapid identification of human ovarian cancer in second harmonic generation images using radiomics feature analyses and tree-based pipeline optimization tool. *J Biophotonics*. 2020;13:e202000050.
20. Chen D, Chen G, Jiang W, et al. Association of the collagen signature in the tumor microenvironment with lymph node metastasis in early gastric cancer. *JAMA Surg*. 2019;154:e185249.
21. Huang Y-Q, Liang C-H, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. 2016;34:2157-2164.
22. Huang X, Wu Z, Zhou A, et al. Nomogram combining radiomics with the American college of radiology thyroid imaging reporting and data system can improve predictive performance for malignant thyroid nodules. *Front Oncol*. 2021;11:737847.
23. Wang S, Chen Y, Zhang H, Liang Z, Bu J. The value of predicting human epidermal growth factor receptor 2 status in adenocarcinoma of the esophagogastric junction on CT-based radiomics nomogram. *Front Oncol*. 2021;11:707686.
24. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016;34:2534-2540.
25. Mouw JK, Ou G, Weaver VM. Extracellular matrix assembly: a multiscale deconstruction. *Nat Rev Mol Cell Biol*. 2014;15:771-785.
26. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol*. 2014;15:786-801.
27. Levental KR, Yu H, Kass L, et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell*. 2009;139:891-906.
28. Zhou ZH, Ji CD, Xiao HL, Zhao HB, Cui YH, Bian XW. Reorganized collagen in the tumor microenvironment of gastric cancer and its association with prognosis. *J Cancer*. 2017;8:1466-1476.
29. Han W, Chen S, Yuan W, et al. Oriented collagen fibers direct tumor cell intravasation. *Proc Natl Acad Sci USA*. 2016;113:11208-11213.

30. Szulczewski JM, Inman DR, Proestaki M, Notbohm J, Burkel BM, Ponik SM. Directional cues in the tumor microenvironment due to cell contraction against aligned collagen fibers. *Acta Biomater*. 2021;129:96-109.

31. Xi G, Guo W, Kang D, et al. Large-scale tumor-associated collagen signatures identify high-risk breast cancer patients. *Theranostics*. 2021;11:3229-3243.

32. Chen W, Dong S, Liu X, et al. Association of the collagen signature in the tumor microenvironment with recurrence and survival of patients with T4N0M0 colon cancer. *Dis Colon Rectum*. 2021;64:563-575.

33. Karimian-Jazi K, Münch P, Alexander A, et al. Monitoring innate immune cell dynamics in the glioma microenvironment by magnetic resonance imaging and multiphoton microscopy (MR-MPM). *Theranostics*. 2020;10:1873-1883.

34. Fu X, Sompol P, Brandon J, et al. In vivo single-molecule detection of nanoparticles for multiphoton fluorescence correlation spectroscopy to quantify cerebral blood flow. *Nano Lett*. 2020;20:6135-6141.

35. Jiang Y, Zhang QI, Hu Y, et al. ImmunoScore signature: a prognostic and predictive tool in gastric cancer. *Ann Surg*. 2018;267:504-513.

36. Meier L, Geer SVD, Bhlmann P, Zrich ETH. The group Lasso for logistic regression. *J R Stat Soc B*. 2008;70:53-71.

37. Jian H, Ma S, Zhang CH. Adaptive LASSO for sparse high-dimensional regression. *Stat Sin*. 2008;18:1603-1618.

38. Grossman M, Ben-Chetrit N, Zhuravlev A, et al. Tumor cell invasion can be blocked by modulators of collagen fibril alignment that control assembly of the extracellular matrix. *Cancer Res*. 2016;76:4249-4258.

39. Martins Cavaco AC, Damaso S, Casimiro S, Costa L. Collagen biology making inroads into prognosis and treatment of cancer progression and metastasis. *Cancer Metastasis Rev*. 2020;39:603-623.

40. Provenzano PP, Inman DR, Eliceiri KW, et al. Collagen density promotes mammary tumor initiation and progression. *BMC Med*. 2008;6:11.

41. Xu S, Xu H, Wang W, et al. The role of collagen in cancer: from bench to bedside. *J Transl Med*. 2019;17:309.

42. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. *Am J Neuroradiol*. 2010;31:809-816.

43. Xu L, Yang P, Liang W, et al. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. *Theranostics*. 2019;9:5374-5385.

44. Wyckoff JB, Wang Y, Lin EY, et al. Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. *Cancer Res*. 2007;67:2649-2656.

45. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer*. 2013;13:97-110.

46. Fang M, Yuan J, Peng C, Li Y. Collagen as a double-edged sword in tumor progression. *Tumour Biol*. 2014;35:2871-2882.

47. Wei SC, Fattet L, Tsai JH, et al. Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nat Cell Biol*. 2015;17:678-688.

48. Kumar S, Das A, Sen S. Extracellular matrix density promotes EMT by weakening cell-cell adhesions. *Mol Biosyst*. 2014;10:838-850.

49. Zhang K, Corsa CA, Ponik SM, et al. The collagen receptor discoidin domain receptor 2 stabilizes SNAIL1 to facilitate breast cancer metastasis. *Nat Cell Biol*. 2013;15:677-687.

50. Koenig A, Mueller C, Hasel C, Adler G, Menke A. Collagen type I induces disruption of E-cadherin-mediated cell-cell contacts and promotes proliferation of pancreatic carcinoma cells. *Cancer Res*. 2006;66:4662-4671.

51. Imamichi Y, König A, Gress T, Menke A. Collagen type I-induced Smad-interacting protein 1 expression downregulates E-cadherin in pancreatic cancer. *Oncogene*. 2007;26:2381-2385.

52. Nie KE, Shi L, Chen Q, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res*. 2016;22:5256-5264.

53. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol Med*. 2021;126:1296-1311.

54. Ueno H, Ishiguro M, Nakatani E, et al. Prognostic value of desmoplastic reaction characterisation in stage II colon cancer: prospective validation in a Phase 3 study (SACURA Trial). *Br J Cancer*. 2021;124:1088-1097.

55. Ueno H, Kajiwara Y, Ajioka Y, et al. Histopathological atlas of desmoplastic reaction characterization in colorectal cancer. *Jpn J Clin Oncol*. 2021;51:1004-1012.

56. Blockhuys S, Van Rompaye B, De Rycke R, et al. Radiation-induced myosin IIA expression stimulates collagen type I matrix reorganization. *Radiother Oncol*. 2013;108:162-167.

57. Mills BN, Qiu H, Drage MG, et al. Modulation of the human pancreatic ductal adenocarcinoma immune microenvironment by stereotactic body radiotherapy. *Clin Cancer Res*. 2022;28:150-162.

58. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis*. 1997;12:19-23.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.