iScience

Article

Development and validation of a collagen signature to predict the prognosis of patients with stage II/III colorectal cancer

Shumin Dong,
Huaiming Wang,
Hongli Ji, ...,
Shuangmu Zhuo,
Gang Chen, Jun
Yan

shuangmuzhuo@gmail.com
(S.Z.)
naichengang@126.com (G.C.)
yanjunfudan@163.com (J.Y.)

Highlights
A TME collagen signature
(CS$_{TME}$) of CRC was
developed and validated

The CS$_{TME}$ was an
independent prognostic
risk factor

The "TNM stage+CS$_{TME}$"
had a better prognostic
value than that of TNM
stage alone

Article

Development and validation of a collagen signature to predict the prognosis of patients with stage II/III colorectal cancer

Shumin Dong,[1,2,8] Huaiming Wang,[3,4,8] Hongli Ji,[1,8] Yaowen Hu,[1] Shuhan Zhao,[1] Botao Yan,[1] Guangxing Wang,[2,5] Zexi Lin,[6] Weifeng Zhu,[7] Jianping Lu,[7] Jiaxin Cheng,[1] Zhida Wu,[7] Qiong Zhu,[7] Shuangmu Zhuo,[2,*] Gang Chen,[7,*] and Jun Yan[1,9,*]

SUMMARY

The tumor, nodes and metastasis (TNM) classification system provides useful but incomplete prognostic information and lacks the assessment of the tumor microenvironment (TME). Collagen, the main component of the TME extracellular matrix, plays a nonnegligible role in tumor invasion and metastasis. In this cohort study, we aimed to develop and validate a TME collagen signature (CS$_{TME}$) for prognostic prediction of stage II/III colorectal cancer (CRC) and to compare the prognostic values of "TNM stage + CS$_{TME}$" with that of TNM stage alone. Results indicated that the CS$_{TME}$ was an independent prognostic risk factor for stage II/III CRC (hazard ratio: 2.939, 95% CI: 2.180–3.962, p < 0.0001), and the integration of the TNM stage and CS$_{TME}$ had a better prognostic value than that of the TNM stage alone (AUC$_{(TNM+CSTME)}$ = 0.772, AUC $_{TNM}$ = 0.687, p < 0.0001). This study provided an application of "seed and soil" strategy for prognosis prediction and individualized therapy.

INTRODUCTION

Colorectal cancer (CRC) is the second most commonly occurring cancer in men, and the third most common cancer diagnosed in women worldwide.[1] Thirty-five percent of CRC patients are diagnosed at stage II/III,[2] and their outcomes differ even if the patients are in the same stage and receive the same treatment.[3] The TNM classification system[4] provides useful prognostic information, but the information is inadequate for precision individualized prognosis prediction,[5] and thus, new supplementary prognostic strategies are required.

In the "seed and soil" theory proposed by Stephen Paget, tumor cells play the role of the "seed" and the tumor microenvironment (TME) is the "soil".[6] The TNM classification system mainly focuses on the "seed" and lacks the assessment of the primary tumor "soil". Although the TME, as the primary tumor "soil", has been reported to be an important factor in tumor progression and metastasis,[7] TME-based prognostic strategies have rarely been investigated.

As one of the main components of the TME extracellular matrix,[8] collagen shapes the TME structure and plays a nonnegligible role in tumor progression, invasion, and metastasis.[9,10] Growing evidence has highlighted the importance of collagen in CRC progression and metastasis. Collagen deposition and modification have been found during CRC progression,[11] and collagen has been reported to play a supportive role in CRC cell invasion and metastasis[12,13] and has been regarded as a potential biomarker.[14] Therefore, we hypothesized that microenvironmental collagen features in the TME ("soil") of stage II/III CRC were associated with the prognosis and the combination of the TNM stage and TME collagen signature (CS$_{TME}$) could have a better prognostic value than that of the TNM stage alone.

The purpose of this study was to investigate whether the CS$_{TME}$ is a prognostic factor in predicting the prognosis of stage II/III CRC and to compare prognostic values of "TNM stage + CS$_{TME}$" and the TNM stage alone.

[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 510515, China

[2]School of Science, Jimei University, Xiamen 361021, China

[3]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

[4]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases Supported by National Key Clinical Discipline, Guangzhou 510630, China

[5]Center for Molecular Imaging and Translational Medicine, Xiamen University, Xiamen 361021, China

[6]Fujian University, Fuzhou 350000, China

[7]Department of Pathology & Precision Medicine Center, The Affiliated Cancer Hospital of Fujian Medical University, Fujian Provincial Cancer Hospital, Fuzhou 350011, China

[8]These authors contributed equally

[9]Lead contact

*Correspondence:
shuangmuzhuo@gmail.com (S.Z.),
naichengang@126.com (G.C.),
yanjunfudan@163.com (J.Y.)
https://doi.org/10.1016/j.isci.2023.106746

RESULTS

Patients

The study comprised 570 patients, including 344 in the training set (mean [SD] age: 57.7[13.9] years, including 185 men [53.8%]) and 226 in the validation set (mean [SD] age: 58.1[12.7] years, including 136 men [60.2%]). The demographic and clinicopathological characteristics were similar in the two sets. The 5-year disease-free survival (DFS) rate was 69.8% (240/344) in the training set and 67.8% (153/226) in the validation set. The 5-year overall survival (OS) rate was 73.5% (253/344) in the training set and 69.9% (158/226) in the validation set. The mean follow-up time was 74.3 months in the training set and 59.8 months in the validation set. The median follow-up time was 86 months in the training set and 67 months in the validation set.

Collagen features in the TME

Collagen features of the center of the tumor (CT), invasive margin (IM) and N regions in the TME were analyzed. Structural damage, reduced components, crosslink rarefaction, and texture loss were found in the CT region. Most features were unidirectionally changed from the CT to IM and then to the N region, while some features, such as the collagen fiber length and gray-level co-occurrence matrix (GCLM) energy, reached anomalous peaks in the IM region. Therefore, interregional variation (CT-IM)/(IM-N) was calculated and further analyzed.

Collagen signature establishment

A total of 994 collagen features in the training set were classified into high and low categories according to 5-year DFS in the training set, and the features in the validation set were classified into high and low categories based on their cutoff values in the training set. Then, 18 features were selected by Least-absolute shrinkage and selection operator (LASSO) Cox regression analysis to establish the CS_{TME} (Figure 1, Table S3).

The CS_{TME} was an independent prognostic risk factor

The CS_{TME} cutoff point (2.527) was determined based on the 5-year DFS in the training set and was used to classify patients into high-CS_{TME} and low-CS_{TME} groups. The clinicopathological characteristics of the high-CS_{TME} and low-CS_{TME} groups were similar (Table S4). Patients with high CS_{TME} presented with a poor prognosis. The 5-year DFS rate was 47.1% (81/172) in the high-CS_{TME} group and 78.4% (312/398) in the low-CS_{TME} group (p < 0.0001). Consistent with this, the 5-year OS rate of the high-CS_{TME} group (50.6%, 87/172) was significantly lower than that of the low-CS_{TME} group (81.4%, 324/398, p < 0.0001).

Univariate Cox regression analysis of clinicopathological characteristics and the CS_{TME} was performed, as shown in Table 1. The depth of invasion, lymph node metastasis status, CS_{TME}, elevated carcinoembryonic antigen (CEA) level, elevated carbohydrate antigen 19-9 (CA19-9) level, and VELIPI (VE: vascular emboli, LI: lymphatic invasion, PI: perineural invasion) status were found to be independent risk factors for 5-year DFS in both the training and validation sets. The depth of invasion, lymph node metastasis, and CS_{TME} remained powerful factors in the multivariate Cox regression analysis (Table 2), while the CEA level, CA19-9 level, and VELIPI status did not. The CS_{TME} remained an independent risk factor for 5-year OS in both the training and validation sets (Tables S5 and S6). Patients with high CS_{TME} had a poor prognosis; the hazard ratio (HR) of the CS_{TME} (high vs. low) for predicting 5-year DFS was 2.939 (95% CI: 2.180–3.962, p < 0.0001), and the HR for predicting 5-year OS was 2.928 (95% CI: 2.138–4.011, p < 0.0001) (Tables S7 and S8).

The CS_{TME} was a powerful tool to discriminate and predict prognosis

Kaplan-Meier survival analyses were conducted between the high-CS_{TME} and low-CS_{TME} groups and the low-CS_{TME} group had a significantly longer 5-year DFS (Figure 2A). The estimated mean 5-year DFS of the low-CS_{TME} group was 14.0 months longer than that of the high-CS_{TME} group (51.9 months vs. 37.9 months, p < 0.0001). The 5-year DFS of the patients with low CS_{TME} was 16.8 months longer in the training set (53.2 months vs. 36.4 months, p < 0.0001) and 9.6 months longer in the validation set (49.8 months vs. 40.2 months, p = 0.0024) than those of the patients with high CS_{TME} (Figures 2B and 2C). In line with this, the low-CS_{TME} group had a significantly longer 5-year OS than that of the high-CS_{TME} group (Figures S2A–S2C). The estimated mean 5-year OS of the patients with low CS_{TME} was 8.6 months longer than that of the patients with high CS_{TME} (54.3 months vs. 45.7 months, p < 0.0001). The CS_{TME}
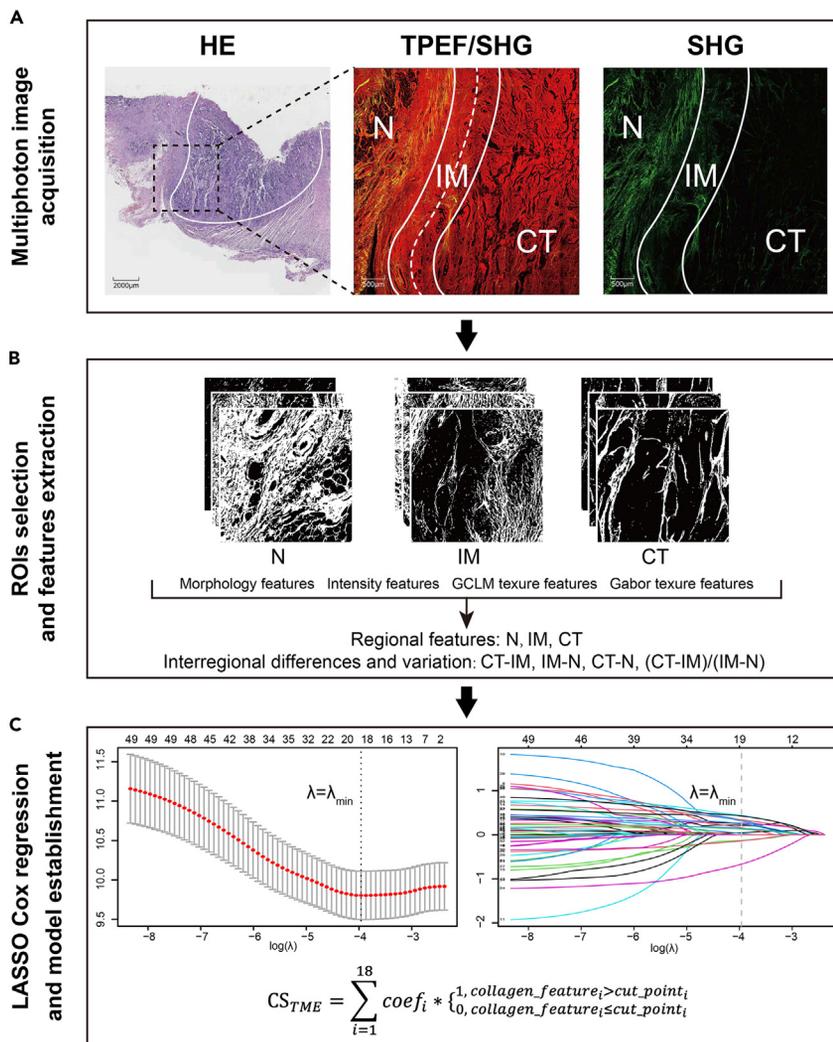
Figure 1. Collagen Signature Establishment

(A) Representative acquired multiphoton images and region divisions. TPEF: 2-photon excitation fluorescence; SHG: second harmonic generation.

(B) ROIs selection and features extraction.

(C) LASSO Cox regression and model establishment. The tuning parameter (λ) was set at λ_{min}, and 18 optimal features were selected against the log (λ) sequence.

remained powerful in distinguishing the prognosis in clinicopathological subgroups (Figures S3 and S4; Table S9).

Time-dependent receiver operating characteristic (ROC) analysis was performed in both the training and validation sets to assess the prognostic value of the CS_{TME} in predicting 5-year DFS and 5-year OS. The CS_{TME} showed good discrimination of 5-year DFS with an AUC of 0.718 (95% CI: 0.674–0.762); the sensitivity was 67.8% and the specificity was 69.0% at the optimal cutoff point (Figure 2D). The AUC of the CS_{TME} for predicting 5-year DFS was 0.729 (95% CI: 0.671–0.787) in the training set and 0.702 (95% CI: 0.632–0.772) in the validation set (Figures 2E and 2F). The CS_{TME} retained its approximative capacity for predicting 5-year OS, with an AUC of 0.720 (95% CI: 0.671–0.787) (Figures S2D–S2F).

"TNM stage+CS_{TME}" had a better prognostic value than that of TNM stage alone

Time-dependent ROC analysis was performed to compare the prognostic value of "TNM stage + CS_{TME}" and the TNM stage alone for predicting 5-year DFS and 5-year OS. The AUC of "TNM stage + CS_{TME}" for

Table 1. Univariate analysis of the CS$_{TME}$ and clinicopathological features in predicting 5-year DFS

Variables	Training Set				Validation Set			
	HR	95% CI		p value	HR	95% CI		p value
Sex (male vs. female)	0.858	0.584	1.261	0.44	0.997	0.624	1.593	0.99
Age (> 50 vs. ≤ 50 years)	1.322	0.853	2.049	0.21	1.369	0.812	2.310	0.24
Depth of Invasion (T4 vs. T1+T2+T3)	2.403	1.410	4.097	0.0013	2.379	1.219	4.641	0.011
Lymph Node Metastasis Status (yes vs. no)	2.326	1.563	3.461	< 0.0001	3.501	2.107	5.819	< 0.0001
Primary Location (colon vs. rectum)	0.931	0.721	1.202	0.58	0.986	0.736	1.321	0.92
Differentiation Status (poor+undifferented vs. well+moderate)	1.576	0.948	2.620	0.079	1.611	0.913	2.843	0.10
Histologic Type (others vs. adenocarcinoma)	1.418	0.658	3.053	0.37	1.695	0.777	3.697	0.18
VELIPI Status (yes vs. no)	2.057	1.303	3.248	0.0020	2.218	1.314	3.744	0.0029
CEA Level (elevated vs. normal)	1.647	1.109	2.446	0.013	1.647	1.027	2.640	0.038
CA19-9 Level (elevated vs. normal)	1.639	1.015	2.645	0.043	1.776	1.069	2.948	0.026
Adjuvant Chemotherapy History (yes vs. no)	0.719	0.489	1.056	0.093	0.860	0.543	1.362	0.52
CS$_{TME}$ (high vs. low)	4.003	2.711	5.911	< 0.0001	2.183	1.374	3.469	0.00095

predicting 5-year DFS was 0.772 (95% CI: 0.730–0.815) and was significantly higher than that of the TNM stage alone (AUC$_{TNM}$ = 0.687, 95% CI: 0.642–0.732, p < 0.0001) (Figures 3A–3C). The sensitivity and specificity of the TNM stage at the optimal cutoff point were 64.4% and 69.5%, respectively, while those of "TNM stage + CS$_{TME}$" were 73.4% and 71.8%, respectively. Consistently, the AUC of the "TNM stage + CS$_{TME}$" combination in predicting 5-year OS was 0.780 (95% CI: 0.736–0.823) and was significantly higher than that of the TNM stage alone (AUC$_{TNM}$ = 0.698, 95% CI: 0.651–0.745, p < 0.0001) (Figures S2G–S2I).

According to the multivariate Cox regression analysis, we established a "TNM stage + CS$_{TME}$" nomogram, which included the depth of invasion, lymph node metastasis status and CS$_{TME}$ (Figure 3D). Calibration curve and discrimination assessments were performed using the concordance index. The analysis revealed fair calibration of the nomogram, which was demonstrated to be clinically useful by decision curve analysis (Figures 3E and 3F).

Low-CS$_{TME}$ patients benefited from adjuvant chemotherapy

Kaplan-Meier survival analyses of the patients treated and untreated with adjuvant chemotherapy were performed, and the results indicated that low-CS$_{TME}$ patients benefited from adjuvant chemotherapy rather than high-CS$_{TME}$ patients. Adjuvant chemotherapy helped the patients with low CS$_{TME}$ gain 6.5 months in the estimated mean 5-year DFS (54.7 months vs. 48.2 months, p = 0.00035) and 5.4 months in the estimated mean 5-year OS (56.7 months vs. 51.3 months, p = 0.00015), but there was no significant effect in the patients with high CS$_{TME}$ (5-year DFS: 36.9 months vs. 39.4 months, p = 0.47; 5-year OS: 46.0 months vs. 45.4 months, p = 0.84).

Table 2. Multivariate analysis of the CS$_{TME}$ and clinicopathological features in predicting 5-year DFS

Variables	Training Set				Validation Set			
	HR	95% CI		p value	HR	95% CI		p value
Depth of Invasion (T4 vs. T1+T2+T3)	2.156	1.207	3.850	0.0094	2.370	1.169	4.804	0.017
Lymph Node Metastasis Status (yes vs. no)	1.937	1.275	2.944	0.0019	3.373	1.938	5.869	< 0.0001
VELIPI Status (yes vs. no)	1.294	0.778	2.153	0.32	1.165	0.663	2.045	0.60
CEA Level (elevated vs. normal)	1.445	0.964	2.165	0.075	1.056	0.642	1.737	0.83
CA19-9 Level (elevated vs. normal)	1.168	0.712	1.917	0.54	1.719	0.988	2.990	0.055
CS$_{TME}$ (high vs. low)	3.749	2.513	5.593	< 0.0001	2.288	1.407	3.718	0.00084
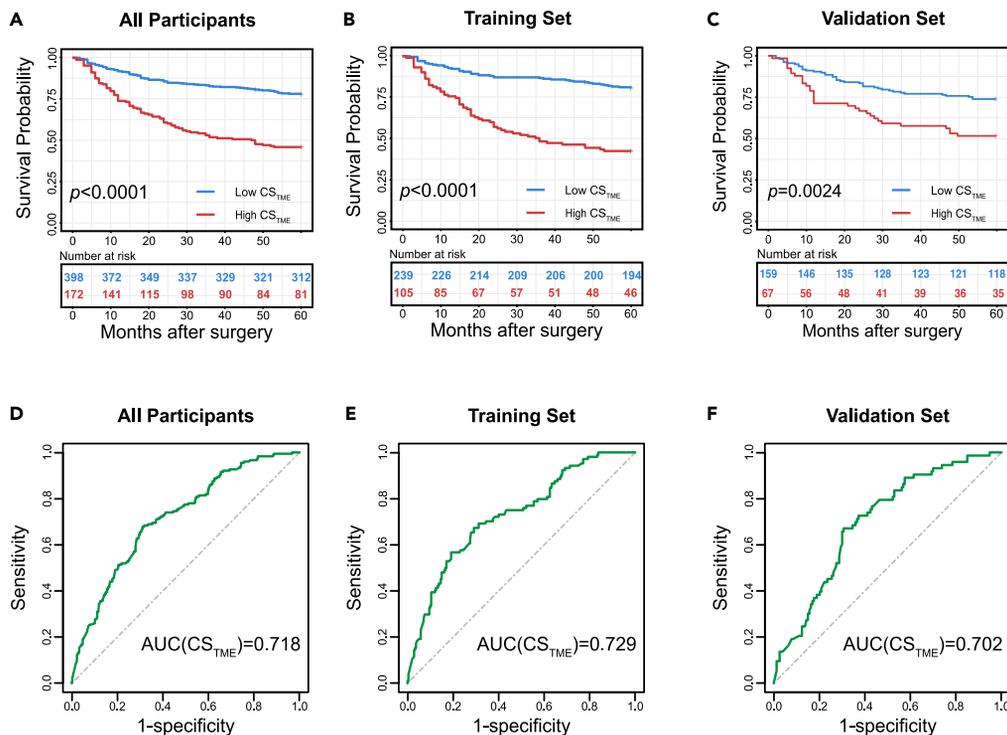
Figure 2. The CS$_{TME}$ was a powerful tool to discriminate and predict prognosis
(A) Kaplan-Meier survival analysis of the high-CS$_{TME}$ and low-CS$_{TME}$ groups for predicting 5-year DFS in all patients.
(B) Kaplan-Meier survival analysis of the high-CS$_{TME}$ and low-CS$_{TME}$ groups for predicting 5-year DFS in the training set.
(C) Kaplan-Meier survival analysis of the high-CS$_{TME}$ and low-CS$_{TME}$ groups for predicting 5-year DFS in the validation set.
(D) Time-dependent ROC curve of the CS$_{TME}$ for predicting 5-year DFS in all patients.
(E) Time-dependent ROC curve of the CS$_{TME}$ for predicting 5-year DFS in the training set.
(F) Time-dependent ROC curve of the CS$_{TME}$ for predicting 5-year DFS in the validation set.

DISCUSSION

In this cohort study, a CS$_{TME}$ was developed and validated for predicting the prognosis of stage II/III CRC. The results showed that the CS$_{TME}$ was a prognostic risk factor of stage II/III CRC, and patients with high CS$_{TME}$ had poor outcomes. The integration of "TNM stage + CS$_{TME}$" led to a better prognostic value than that of the TNM stage alone. Thus, the CS$_{TME}$ could serve as a new prognostic tool and provide a reference for clinical practice.

The prognosis of stage II/III CRC is heterogeneous. The TNM classification system provides useful but incomplete prognostic information.[5] TNM classification mainly focuses on the "seed" feature of tumors and lacks the assessment of the TME. Collagen is one of the main components of the TME and has been reported to play a supportive role in CRC progression and metastasis.[12,13] In this study, we found that the fiber area, fiber number, and fiber length in the IM region; fiber orientation in the N region; fiber crosslink density in the CT region/minor IM region; and several texture features were associated with a poor prognosis. Then, we established the CS$_{TME}$ and an integrated "TNM stage + CS$_{TME}$" prognostic nomogram for patients diagnosed with stage II/III CRC. The CS$_{TME}$ was revealed to be an independent prognostic risk factor. "TNM stage + CS$_{TME}$" had a better prognostic value than that of the TNM stage alone. The "TNM stage + CS$_{TME}$" nomogram showed fair calibration and was demonstrated to be clinically useful by decision curve analysis, providing an individualized prognostic tool for patient stratification.

Multiphoton imaging has been widely used in real-time *in vivo* imaging and optical biopsy and provides a powerful modality for detecting collagen in diverse tissues.[15] By analyzing morphological and textural collagen features extracted from images, multidimensional quantitative metrics can be employed to diagnose and predict diseases.[16–20] However, many of the studies to date have unsatisfactorily evaluated the
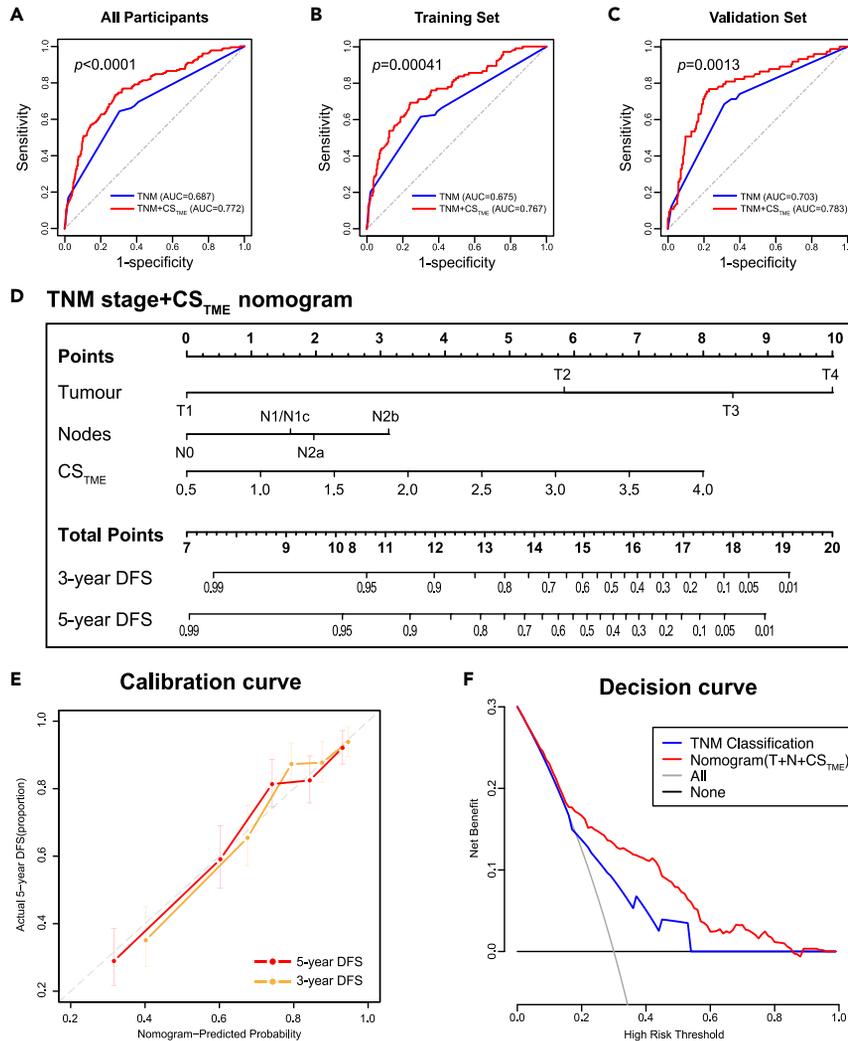
Figure 3. "TNM stage+CS_TME" had a better prognostic value than that of TNM stage alone
(A) Time-dependent ROC curves of the TNM stage and "TNM stage + CS_TME" for predicting 5-year DFS in all patients.
(B) Time-dependent ROC curves of the TNM stage and "TNM stage + CS_TME" for predicting 5-year DFS in the training set.
(C) Time-dependent ROC curves of the TNM stage and "TNM stage + CS_TME" for predicting 5-year DFS in the validation set.
(D) The "TNM stage + CS_TME" nomogram.
(E) Calibration curves of 3-year DFS and 5-year DFS.
(F) Decision curves of the TNM stage and "TNM stage + CS_TME" nomogram.

heterogeneity and internal variation of the TME. In this study, to obtain a comprehensive assessment of the CS_TME, multiphoton images were divided into CT, IM and N regions, and interregional differences and variation were further calculated. All the features were used to establish the CS_TME. Previous studies have indicated that collagen plays a role as a "highway" for tumor invasion and metastasis.[9,21] In our study, the CS_TME included the collagen fiber area, fiber number, and fiber length in the IM area, the specific TME region where collagen acts as a "highway", as prognostic risk parameters. The CS_TME also considered the fiber crosslink density of the CT region/minor IM region, which describes the interregional variance of the fiber crosslink density between the CT and IM regions, as a protective parameter, suggesting that lower crosslink extrusion might provide a sparse "highway network" for tumor cells to metastasis. In addition, the N fiber orientation was selected as a protective prognostic parameter in the CS_TME. This feature describes the isotropy of collagen fibers, indicating that N regions with anisotropic collagen fibers provide less support for tumor invasion than those with isotropic collagen fibers. Overall, high CS_TME were associated with a poor prognosis, and the morphological collagen features of the TME in patients with high CS_TME were
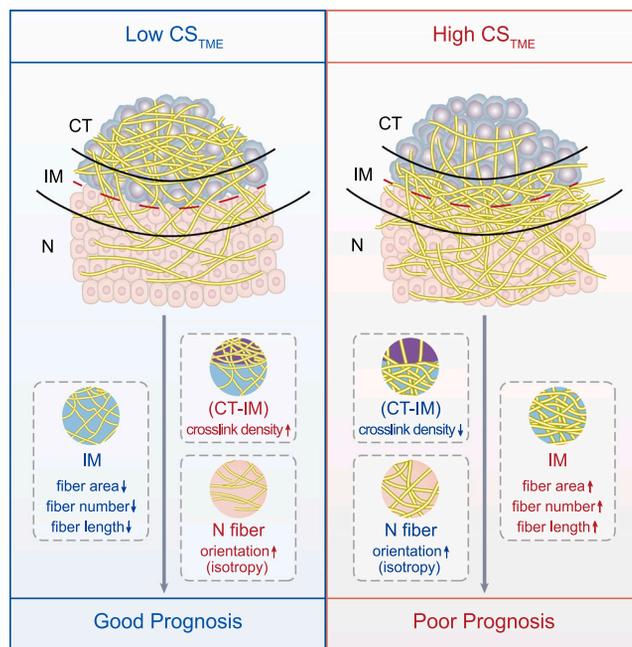
Figure 4. TME morphological features of the high-CS$_{TME}$ and low-CS$_{TME}$ groups

characterized by increased fiber area, number, and length in the IM region, decreased fiber orientation in the N region, and a decreased difference in the fiber crosslink density between the CT and IM regions (as shown in Figure 4).

The identification of subgroups of patients with CRC who might benefit from adjuvant chemotherapy is challenging.[22] Our study found that low-CS$_{TME}$ patients benefited from adjuvant chemotherapy rather than high-CS$_{TME}$ patients. Patients with low CS$_{TME}$ who underwent adjuvant chemotherapy had significantly longer estimated mean 5-year DFS and 5-year OS than patients who did not receive adjuvant chemotherapy. Meanwhile, there was no significant effect of adjuvant chemotherapy in patients with high CS$_{TME}$, which suggests that the CS$_{TME}$ is a new classification tool for clinical decision-making. Data from published sources have indicated that the increased collagen crosslinking and content are associated with reduced drug delivery into tumor cells and lead to chemoresistance.[23,24] Moreover, previous studies have reported that collagen can promote treatment escape and the development of early micrometastases.[25,26] The lack of a significant beneficial effect of adjuvant chemotherapy in the high-CS$_{TME}$ group in the survival analyses can possibly be attributed to the collagen-related inefficacy of drug delivery, as well as collagen-associated tumor cell escape and invasion after chemotherapy.

The current study innovatively and comprehensively assessed the TME collagen features of patients with stage II/III CRC, and established a TME-based prognostic tool (CS$_{TME}$). By considering both the features of the TME ("soil") and the tumor ("seed"), we generated a "TNM stage + CS$_{TME}$" prognostic system for stage II/III CRC. The results indicated that "TNM stage + CS$_{TME}$" had a better prognostic value than that of the TNM stage alone, and the "TNM stage + CS$_{TME}$" nomogram was demonstrated to be clinically useful. The "TNM stage + CS$_{TME}$" strategy can provide complementary prognostic information for individualized therapy for stage II/III CRC.

Conclusions

The CS$_{TME}$ is an independent and effective prognosticator in patients with stage II/III CRC, and the combination of the TNM stage and CS$_{TME}$ has a better prognostic value than that of the TNM stage alone.

Limitations of the study

A limitation of the study is that it was retrospective, and there might have been potential selection bias. Future studies on the generalizability of the data could be performed in multicenter validations of diverse

regions and ethnicities. The CS_{TME} mainly assessed the structural signature of TME extracellular matrix by detecting collagens with triple helix structure. Further studies, which take other TME contents into account, will need to be undertaken.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Study design
  - ○ Multiphoton imaging and region of interest (ROI) selection
  - ○ Collagen feature extraction
  - ○ CS_{TME} development and validation
  - ○ Prognostic value assessment and nomogram construction
  - ○ Outcomes
- QUANTIFICATION ANTD STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106746.

AUTHOR CONTRIBUTIONS

S.D., H.W., and J.Y. conceived and designed this study. S.D., H.W., S.H.Z., and B.Y. contributed to patient recruitment and clinicopathological data collection. S.M.Z. coordinated all multiphoton laboratory aspects of the project. S.D., G.W., Z.L., and H.J. performed multiphoton imaging. G.C. coordinated all pathology-related work. W.Z., J.C., and Y.H. were involved in hematoxylin and eosin image acquisition. J.L., Z.W., Q.Z., and G.C. participated in image division and region of interest (ROI) selection. Extraction of the collagen features was performed by S.H.Z. and B.Y. Data analysis and result interpretation were performed by S.D. H.J. completed the clinical trial registration and ethical procedures. S.D. and J.Y. drafted the manuscript. H.W., S.M.Z., and G.C. critically revised the manuscript. All authors approved the final draft submitted. S.D. and H.W. contributed equally as first authors. S.M.Z., G.C., and J.Y. contributed equally as corresponding authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *71*, 209–249. https://doi.org/10.3322/CAAC.21660.

2. Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Cercek, A., Smith, R.A., and Jemal, A. (2020). Colorectal cancer statistics, 2020. CA Cancer J. Clin. *70*, 145–164. https://doi.org/10.3322/caac.21601.

3. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2021). Cancer statistics, 2021. CA Cancer J. Clin. *71*, 7–33. https://doi.org/10.3322/caac.21654.

4. M.B. Amin, S.B. Edge, F.L. Greene, D.R. Byrd, R.K. Brookland, M.K. Washington, J.E. Gershenwald, C.C. Compton, K.R. Hess, and D.C. Sullivan, et al., eds. (2017). AJCC Cancer Staging Manual, Eighth edition (Springer International Publishing). https://doi.org/10.1007/978-3-319-40618-3.

5. O'Sullivan, B., Brierley, J., Byrd, D., Bosman, F., Kehoe, S., Kossary, C., Piñeros, M., Van Eycken, E., Weir, H.K., and Gospodarowicz, M. (2017). The TNM classification of malignant tumours-towards common understanding and reasonable expectations. Lancet Oncol. *18*, 849–851. https://doi.org/10.1016/S1470-2045(17)30438-2.

6. Witz, I.P., and Levy-Nissenbaum, O. (2006). The tumor microenvironment in the post-PAGET era. Cancer Lett. *242*, 1–10. https://doi.org/10.1016/j.canlet.2005.12.005.

7. Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. Nat. Med. *19*, 1423–1437. https://doi.org/10.1038/NM.3394.

8. Theocharis, A.D., Skandalis, S.S., Gialeli, C., and Karamanos, N.K. (2016). Extracellular matrix structure. Adv. Drug Deliv. Rev. *97*, 4–27. https://doi.org/10.1016/j.addr.2015.11.001.

9. Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T., Zhang, W., Yu, X., and Liu, L. (2019). The role of collagen in cancer: from bench to bedside. J. Transl. Med. *17*, 309. https://doi.org/10.1186/s12967-019-2058-1.

10. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat. Commun. *11*, 5120. https://doi.org/10.1038/S41467-020-18794-X.

11. Le, C.C., Bennasroune, A., Langlois, B., Salesse, S., Boulagnon-Rombi, C., Morjani, H., Dedieu, S., and Appert-Collin, A. (2020). Functional interplay between collagen network and cell behavior within tumor microenvironment in colorectal cancer. Front. Oncol. *10*, 527. https://doi.org/10.3389/fonc.2020.00527.

12. Vellinga, T.T., Den Uil, S., Rinkes, I.H.B., Marvin, D., Ponsioen, B., Alvarez-Varela, A., Fatrai, S., Scheele, C., Zwijnenburg, D.A., Snippert, H., et al. (2016). Collagen-rich stroma in aggressive colon tumors induces mesenchymal gene expression and tumor cell invasion. Oncogene *35*, 5263–5271. https://doi.org/10.1038/ONC.2016.60.

13. Jeitany, M., Leroy, C., Tosti, P., Lafitte, M., Le Guet, J., Simon, V., Bonenfant, D., Robert, B., Grillet, F., Mollevi, C., et al. (2018). Inhibition of DDR1-BCR signalling by nilotinib as a new therapeutic strategy for metastatic colorectal cancer. EMBO Mol. Med. *10*, e7918. https://doi.org/10.15252/EMMM.201707918.

14. Kim, M.S., Ha, S.E., Wu, M., Zogg, H., Ronkon, C.F., Lee, M.Y., and Ro, S. (2021). Extracellular matrix biomarkers in colorectal cancer. Int. J. Mol. Sci. *22*, 9185. https://doi.org/10.3390/IJMS22179185.

15. Zipfel, W.R., Williams, R.M., and Webb, W.W. (2003). Nonlinear magic: multiphoton microscopy in the biosciences. Nat. Biotechnol. *21*, 1369–1377. https://doi.org/10.1038/NBT899.

16. Campagnola, P. (2011). Second harmonic generation imaging microscopy: applications to diseases diagnostics. Anal. Chem. *83*, 3224–3231. https://doi.org/10.1021/AC1032325.

17. Chen, D., Liu, Z., Liu, W., Fu, M., Jiang, W., Xu, S., Wang, G., Chen, F., Lu, J., Chen, H., et al. (2021). Predicting postoperative peritoneal metastasis in gastric cancer with serosal invasion using a collagen nomogram. Nat. Commun. *12*, 179. https://doi.org/10.1038/s41467-020-20429-0.

18. Dong, X., Huang, Y., Yu, X., Huang, M., Jiang, W., Chen, D., Wang, G., Zhuo, S., Chi, P., and Yan, J. (2022). Collagen score in the tumor microenvironment predicts the prognosis of rectal cancer patients after neoadjuvant chemoradiotherapy. Radiother. Oncol. *167*, 99–108. https://doi.org/10.1016/J.RADONC.2021.12.023.

19. Xi, G., Qiu, L., Xu, S., Guo, W., Fu, F., Kang, D., Zheng, L., He, J., Zhang, Q., Li, L., et al. (2021). Computer-assisted quantification of tumor-associated collagen signatures to improve the prognosis prediction of breast cancer. BMC Med. *19*, 273. https://doi.org/10.1186/s12916-021-02146-7.

20. Chen, W., Dong, S., Liu, X., Wang, G., Xu, S., Lei, S., Zhuo, S., and Yan, J. (2021). Association of the collagen signature in the tumor microenvironment with recurrence and survival of patients with T4N0M0 colon cancer. Dis. Colon Rectum *64*, 563–575. https://doi.org/10.1097/DCR.0000000000001907.

21. Han, W., Chen, S., Yuan, W., Fan, Q., Tian, J., Wang, X., Chen, L., Zhang, X., Wei, W., Liu, R., et al. (2016). Oriented collagen fibers direct tumor cell intravasation. Proc. Natl. Acad. Sci. USA *113*, 11208–11213. https://doi.org/10.1073/pnas.1610347113.

22. National Cancer Institute (2022). Cancer Stat Facts: Colorectal Cancer (National Cancer Institute).

23. Saatci, O., Kaymak, A., Raza, U., Ersan, P.G., Akbulut, O., Banister, C.E., Sikirzhytski, V., Tokat, U.M., Aykut, G., Ansari, S.A., et al. (2020). Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer. Nat. Commun. *11*, 2416. https://doi.org/10.1038/S41467-020-16199-4.

24. Wang, H., Mislati, R., Ahmed, R., Vincent, P., Nwabunwanne, S.F., Gunn, J.R., Pogue, B.W., and Doyley, M.M. (2019). Elastography can map the local inverse relationship between shear modulus and drug delivery within the pancreatic ductal adenocarcinoma microenvironment. Clin. Cancer Res. *25*, 2136–2143. https://doi.org/10.1158/1078-0432.CCR-18-2684.

25. Pietilä, E.A., Gonzalez-Molina, J., Moyano-Galceran, L., Jamalzadeh, S., Zhang, K., Lehtinen, L., Turunen, S.P., Martins, T.A., Gultekin, O., Lamminen, T., et al. (2021). Co-evolution of matrisome and adaptive adhesion dynamics drives ovarian cancer chemoresistance. Nat. Commun. *12*, 3904. https://doi.org/10.1038/S41467-021-24009-8.

26. Sun, X., He, X., Zhang, Y., Hosaka, K., Andersson, P., Wu, J., Wu, J., Jing, X., Du, Q., Hui, X., et al. (2022). Inflammatory cell-derived CXCL3 promotes pancreatic cancer metastasis through a novel myofibroblast-hijacked cancer escape mechanism. Gut *71*, 129–147. https://doi.org/10.1136/GUTJNL-2020-322744.

27. Zhuo, S., Chen, J., Luo, T., Zou, D., and Zhao, J. (2006). Multimode nonlinear optical imaging of the dermis in ex vivo human skin based on the combination of multichannel mode and Lambda mode. Opt Express *14*,

7810–7820. https://doi.org/10.1364/OE.14.007810.

28. Halama, N., Michel, S., Kloor, M., Zoernig, I., Benner, A., Spille, A., Pommerencke, T., von Knebel, D.M., Folprecht, G., Luber, B., et al. (2011). Localization and density of immune cells in the invasive margin of human colorectal cancer liver metastases are prognostic for response to chemotherapy. Cancer Res. *71*, 5670–5677.

https://doi.org/10.1158/0008-5472.can-11-0268.

29. Xu, S., Wang, Y., Tai, D.C.S., Wang, S., Cheng, C.L., Peng, Q., Yan, J., Chen, Y., Sun, J., Liang, X., et al. (2014). QFibrosis: a fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients.

J. Hepatol. *61*, 260–269. https://doi.org/10.1016/j.jhep.2014.02.015.

30. Xu, S., Kang, C.H., Gou, X., Peng, Q., Yan, J., Zhuo, S., Cheng, C.L., He, Y., Kang, Y., Xia, W., et al. (2016). Quantification of liver fibrosis via second harmonic imaging of the Glisson's capsule from liver surface. J. Biophotonics *9*, 351–363. https://doi.org/10.1002/jbio.201500001.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
R version 4.0.3	R Core Team	https://cran.r-project.org/bin/windows/base/old/4.0.3/
glmnet package in R	Jerome Friedman et al.	https://cran.r-project.org/web/packages/glmnet/index.html
survival package in R	Terry M Therneau et al.	https://cran.r-project.org/web/packages/survival/index.html
survminer package in R	Alboukadel Kassambara et al.	https://cran.r-project.org/web/packages/survminer/index.html
survRM2 package in R	Hajime Uno et al.	https://cran.r-project.org/web/packages/survRM2/index.html
pROC package in R	Xavier Robin et al.	https://cran.r-project.org/web/packages/pROC/index.html
rms package in R	Frank E Harrell Jr	https://cran.r-project.org/web/packages/rms/index.html
Zen version 2.3 lite	Carl Zeiss AG	https://www.zeiss.com/microscopy/en/products/software/zeiss-zen-lite.html
MATLAB 2015b	MathWorks	https://www.mathworks.com/products/matlab.html
Other		
Collagen Features Extraction	Shuoyu Xu et al.	https://doi.org/10.1016/j.jhep.2014.02.015; https://doi.org/10.1002/jbio.201500001
Multiphoton imaging system (Carl Zeiss LSM 880 with Airyscan)	Carl Zeiss AG	https://www.zeiss.com/microscopy

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Jun Yan (yanjunfudan@163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The clinicopathological, survival and collagen data of all participants after deidentification are available from the corresponding author; proposals or written requests for access should be directed to Dr. Jun Yan (yanjunfudan@163.com). The data sharing process should be approved by the Ethics Committee of Nanfang Hospital, and a data access agreement should be signed.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request, Dr. Jun Yan (yanjunfudan@163.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A total of 570 patients diagnosed with stage II/III CRC were enrolled from two independent centers (Figure S1, Table S1). The enrollment of patients was approved by the Ethics Committees of the Nanfang Hospital (NFEC22192) and The Sixth Affiliated Hospital (ZSLYEC22266.5). The training set, including 344 consecutive patients (mean [SD] age: 57.7[13.9] years, including 185 men[53.8%]), was obtained from the Nanfang Hospital in Guangzhou from January 2009 to December 2011. The validation set, comprising 226 consecutive patients (mean [SD] age: 58.1[12.7] years, including 136 men[60.2%]), was obtained from The Sixth Affiliated Hospital of Sun Yat-sen University in Guangzhou from January 2014 to December 2014. The inclusion criteria were as follows: age 18 years or older, pathologically diagnosed with stage II/III CRC (according to the 8th AJCC Cancer Staging Manual), and received radical tumor resection surgery.

The exclusion criteria were as follows: American Society of Anesthesiologists (ASA) grade four, previous treatment with neoadjuvant therapy, hereditary CRC, and pregnancy. The patient enrollment diagram is shown in Figure S1.

The baseline clinical and pathological data of each patient were collected, including sex, age, ASA grade, body mass index, TNM stage, depth of tumor invasion, lymph node metastasis status, primary tumor location, tumor differentiation status, histopathological type, CEA level, CA19-9 level, VELIPI status (VE: vascular emboli, LI: lymphatic invasion, PI: perineural invasion), and adjuvant chemotherapy history. Sex was defined as the biological sex. The age at diagnosis older than 50 years was defined as old. Elevated CEA was defined as a value higher than 5.00 ng/mL (reference: 0.00-5.00 ng/mL), and elevated CA19-9 was defined as a value higher than 37.00 kU/L (reference: 0.00-37.00 kU/L).

All the patients were followed up for five years after surgery with an interval of one year. During follow-up, disease progression and the survival status were recorded. The status of patients was recorded as NED (no evidence of disease), AWD (alive with disease), DOD (died of disease), or DOC (died of other causes). For cases with progression, the date and site of recurrence and/or metastasis were recorded. The death cause and date were recorded for dead cases. DFS was defined as the time from surgery to disease progression or disease-related death, and OS was defined as the time from surgery to death from any cause. Formalin-fixed paraffin-embedded specimens from all patients were used.

METHOD DETAILS

Study design

This study was a retrospective observational cohort study (NCT05420415). The study was approved by the Ethics Committees of the Nanfang Hospital and The Sixth Affiliated Hospital. All procedures involving human participants in this study were in accordance with the Declaration of Helsinki. Written informed consent was waived because of the retrospective design of the study and anonymous data analysis.

Multiphoton imaging and region of interest (ROI) selection

Unstained paraffin tissue sections were imaged with a 20× objective using a multiphoton imaging system.[27] The multiphoton imaging was performed on LSM 880 confocal workstation with Airyscan. The dual-channel imaging mode was used, which independently detects the second harmonic generation (SHG) signals of matrix collagen and autofluorescence signals (TPEF: 2-photon excitation fluorescence) of cells. The SHG signal is excited by the laser with an excitation wavelength of 810 nm and an emission wavelength of 435 nm (Ti: Sapphire Laser Mira-900F, COHERENT INC) to display the intercellular matrix and collagen structure. The TPEF channel detected the autofluorescence signals with wavelength between 430 and 560 nm. The 5 mm × 5 mm imaging area was located in the tumor borderline. The imaging areas were selected by two independent pathologists, and deeper infiltrated sites were preferred. The director of the Pathology Department was consulted if the two pathologists had different opinions.

After multiphoton imaging, the images were divided into three regions: the center of the tumor (CT), invasive margin (IM),[28] and normal tissue (N). Tumor borders were determined manually by two independent pathologists, and the director of the Pathology Department was consulted if the two pathologists had different opinions. The width of the IM area was 1000 μm, which was expanded by 500 μm to the tumor side and the normal tissue side from the tumor borderline. The ZEN software (Zen 2.3 lite, Carl Zeiss Microscopy GmbH, 2011) was used to select ROIs. Three ROIs were randomly selected in the CT, IM, and N regions, the sizes of which were 844 pixels × 844 pixels (700 μm × 700 μm). The prognosis and follow-up data were blinded in the multiphoton image acquisition and ROI selection processes.

Collagen feature extraction

MATLAB 2015b (MathWorks) was used to extract collagen features.[29,30] A total of 142 collagen features in the ROIs were extracted, including eight morphological features, six intensity features, and 128 texture features (Table S2). The mean of the features extracted from the three ROIs in each region was considered the collagen feature of the region. Interregional differences (CT-IM, IM-N, and CT-N) and variation were calculated. Interregional variation was defined as (CT-IM)/(IM-N). Based on the above-mentioned seven groups of the 142 features, a total of 994 features were included and further classified into high and low categories according to the 5-year DFS in the training set.

CS_{TME} development and validation

Least-absolute shrinkage and selection operator (LASSO) Cox regression was performed to select collagen features and establish the CS_{TME}. The penalty parameter λ in this study was selected at minimum (min) criteria by ten-time cross validations. The CS_{TME} was internally validated by bootstrap sampling and externally validated in the validation set.

Prognostic value assessment and nomogram construction

The CS_{TME} was used to classify samples into high and low groups according to 5-year DFS in the training set. Univariate and multivariate analyses were performed. The prognostic value of the CS_{TME} was assessed by Kaplan–Meier survival, receiver operating characteristic (ROC) curve and analyses in both the training and validation sets. The prognostic value of "TNM stage + CS_{TME}" was analyzed by ROC curve analysis and compared with that of the TNM stage alone. A "TNM stage + CS_{TME}" nomogram was constructed according to multivariate analysis and was assessed through calibration curve and decision curve analyses. Adjuvant chemotherapy responses were analyzed in both the high-CS_{TME} and low-CS_{TME} groups by Kaplan–Meier survival analysis (Figure S5).

Outcomes

The primary objective was to investigate whether the CS_{TME} is a prognostic factor in predicting the prognosis of stage II/III CRC. The secondary objective was to compare the prognostic values of "TNM stage + CS_{TME}" and the TNM stage alone. The main outcome was 5-year DFS after surgery, and the secondary outcome was 5-year OS after surgery. The main prognostic value measurement was the area under the ROC curve (AUC) for predicting 5-year DFS, and the secondary prognostic value measurement was the AUC for predicting 5-year OS.

QUANTIFICATION ANTD STATISTICAL ANALYSIS

Feature comparisons of different TME regions were performed by a paired t-test or paired Wilcoxon test. Cross-tabulation analysis was performed with Fisher's exact test. The Kaplan–Meier method was used for survival analysis, and a restricted mean survival time test was performed. The DeLong test was used in the AUC comparison. Cox proportional hazard regression was conducted to compute the hazard ratio (HR). Calibration curve and discrimination assessments were performed using the concordance index.

All statistical analyses were conducted with the R software (version 4.0.3). The R packages "glmnet", "survival", "survminer", "survRM2", "pROC", and "rms" were used. Two-sided $p < 0.05$ was set as the cutoff for a significant difference.

ADDITIONAL RESOURCES

Clinical trial registry: https://clinicaltrials.gov/ct2/show/NCT05420415.