Society of Medical Education & Research

2023, Volume 3, Issue 1, Page No: 63-72

ISSN: 3108-4834

Archive of International Journal of Cancer and Allied Science

Prognostic Value of Collagen Signatures in Breast Cancer: Enhancing Traditional Staging Models

Hendrikus Adrianus van der Meer[1]*, Willem Floris de Vries[1]

[1]Department of Clinical Oncology, Erasmus University Medical Center, Rotterdam, Netherlands.

***E-mail** ✉ h.meer.erasmus@yahoo.com

Abstract

While tumor-associated collagen signature (TACS) is known to independently predict outcomes in breast cancer, it is not yet established whether a full collagen profile—including TACS, TACS-derived microscopic collagen characteristics (TCMF1), and nuclear features associated with TACS (TCMF2)—can enhance the prognostic accuracy of the existing tumor–node–metastasis (TNM) staging system. A total of 941 breast cancer patients were included across three cohorts: training (n = 355), internal validation (n = 334), and external validation (n = 252). Tumor-associated collagen signature (TACS) and TACS-derived microscopic features (TCMF1) were assessed using multiphoton microscopy (MPM), while TACS-derived nuclear features (TCMF2) were extracted from hematoxylin and eosin-stained images aligned with the MPM images. These features were combined linearly to generate a comprehensive collagen signature score, which was then used to refine TNM staging into stage I (II and III)/low-risk and stage I (II and III)/high-risk categories.

Patients with low-risk collagen signatures who were originally classified as stage II or III were effectively "downstaged," while those with stage I tumors and high-risk collagen signatures were "upstaged." Integrating the complete collagen signature into the TNM system markedly improved risk stratification. In the modified staging, stage II–new showed hazard ratios of 8.655, 6.136, and 4.699, and stage III–new had hazard ratios of 14.855, 11.201, and 13.245 across the training, internal, and external validation cohorts, respectively, compared to stage I–new. In contrast, the conventional TNM system exhibited lower hazard ratios for stage II (1.642, 1.853, and 1.371) and stage III (4.131, 4.283, and 3.711) in the same cohorts. Moreover, the modified system outperformed the traditional TNM staging in predictive accuracy, achieving AUCs of 0.843 versus 0.683 in the training cohort, 0.792 versus 0.661 in the internal validation cohort, and 0.793 versus 0.646 in the external validation cohort. The comprehensive collagen signature independently predicts survival outcomes in breast cancer and provides additional insights into the tumor's biological behavior, enhancing the prognostic value of current staging systems.

Keywords: Breast cancer, TNM staging, Collagen signature, Prognosis

Introduction

Breast cancer is the most prevalent malignancy in women and continues to be the leading cause of cancer-related deaths despite major advances in diagnosis and therapy [1]. Reliable staging is essential for evaluating prognosis and guiding treatment selection [2]. Among the clinical staging systems, the tumor–node–metastasis (TNM) framework developed by the American Joint Committee on Cancer (AJCC) is most commonly employed [3]. However, because TNM focuses only on the anatomical extent of disease, it does not capture the underlying biological characteristics of tumors, which can limit its prognostic precision [4]. To address this, the eighth edition of the AJCC incorporated a pathological prognostic staging system, integrating key biomarkers such as estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2, tumor grade, and multigene assays[5, 6] Nonetheless, in regions lacking access to biomarker testing, the conventional TNM system remains the primary staging tool.

Access this article online https://smerpub.com/

Received: 03 November 2022; **Accepted:** 29 January 2023

How to cite this article: van der Meer HA, de Vries WF. Prognostic Value of Collagen Signatures in Breast Cancer: Enhancing Traditional Staging Models. Arch Int J Cancer Allied Sci. 2023;3(1):63-72. https://doi.org/10.51847/dFzZQCviRE

In recent years, the tumor microenvironment has emerged as a source of prognostic biomarkers, among which collagen has received particular attention [7-12]. Multiphoton microscopy (MPM) now allows high-resolution, label-free imaging of multiple tumor microenvironment components on dewaxed tissue sections [13]. Notably, MPM can visualize collagen architecture and its spatial relationship with tumor cells. Tumor-associated collagen signatures (TACSs) describe heterogeneous collagen patterns relative to tumor cells, and TACS-derived microscopic features (TCMF1) have been linked to tumor cell behavior in breast cancer [14-17]. However, it remains uncertain whether combining TACS, TCMF1, and TACS-associated nuclear features (TCMF2) into a comprehensive collagen signature can improve prognostic accuracy beyond the standard TNM staging system.

Nuclear morphology has also been explored as a prognostic marker. For example, Nafe *et al*. [18] demonstrated that parameters such as nuclear size, shape, and spatial relationships correlated with survival in glioblastoma. Similarly, Wang *et al*. [19] applied computational histomorphometric analysis to nuclear orientation, texture, and tumor architecture to predict recurrence in early-stage non-small-cell lung cancer from H&E tissue microarrays. Moreover, the extracellular matrix (ECM) has been shown to influence tumor aggressiveness. Riedl *et al*. [20] reported that tumor cells became more elongated and invasive when moving from dense to porous ECMs. These studies suggest that nuclear features within TACS regions—representing ECM with altered microstructure—could provide additional prognostic information in breast cancer. Integrating TACS, TCMF1, and TCMF2 may therefore enable more precise stratification of patients under the TNM system.

In this study, we applied MPM to capture TACS and extracted both TCMF1 and TCMF2 features to construct a complete collagen signature score (TACS + TCMF1 + TCMF2). Our findings indicate that this comprehensive signature contributes significant prognostic information beyond the TNM system and substantially improves patient stratification. With ongoing development of compact and affordable MPM instruments, this approach has the potential to provide accessible prognostic insights in low- and middle-income countries.

Materials and Methods

Patients

Researchers performed a retrospective review of 1,223 individuals treated surgically for invasive breast carcinoma at two institutions—Fujian Medical University Union Hospital and Harbin Medical University Cancer Hospital—between November 2003 and June 2017. Detailed information regarding eligibility requirements, clinical and pathological features, as well as follow-up details, can be found in the Supplementary Methods section (accessible via https://doi.org/10.1016/j.esmoop.2024.103990).

From this group, 941 cases met the criteria for inclusion in the study. This comprised 689 cases from Fujian Medical University Union Hospital, which were randomly allocated to a training set (n=355) and an internal validation set (n=334), along with an additional 252 cases from Harbin Medical University Cancer Hospital forming the external validation set.

The median age upon diagnosis was 49 years (IQR 42-57) for the training set, 48 years (IQR 40-55) for the internal validation set, and 50 years (IQR 44-57) for the external validation set.

Ethical approval for this research was granted by the institutional review boards at both participating hospitals: Fujian Medical University Union Hospital and Harbin Medical University Cancer Hospital.

Multiphoton imaging and sample preparation system

From formalin-fixed, paraffin-embedded tissue specimens, two consecutive 5-μm sections were prepared. One section was stained with hematoxylin and eosin (H&E), and whole-slide images were captured at ×40 magnification using a VM1000 digital scanner (Motic, Xiamen, China). The second section, which was deparaffinized but left unstained, was imaged using multiphoton microscopy (MPM). The MPM setup was based on a Zeiss LSM 880 laser scanning microscope (Oberkochen, Germany) equipped with a mode-locked femtosecond Ti:Sapphire laser emitting linearly polarized light at 810 nm. Signals from the tissue were collected through two independent channels simultaneously: second harmonic generation signals (395–415 nm) and two-photon excitation fluorescence signals (428–695 nm). Initial large-field images were obtained using a Plan-Apochromat ×10 objective (numerical aperture = 0.45; Zeiss), after which regions of interest were imaged at higher resolution with a Plan-Apochromat ×20 objective (numerical aperture = 0.8; Zeiss).

Generation of TACS-related signatures

The procedures for calculating the macroscopic TACS score and the microscopic TCMF1 score have been described in prior studies [16, 17]. In short, 7–20 discrete, non-overlapping regions (~2.8 mm × 2.8 mm) were selected within and along the edges of the tumor nests on H&E-stained slides. These regions were then mapped to corresponding areas on unstained sections using MPM to identify TACS patterns 1 through 8. Three independent reviewers, blinded to clinical outcomes, confirmed the TACS classifications. The TACS score was then calculated by weighting the frequencies of these eight patterns according to coefficients obtained from ridge regression [16]

For the TCMF1 score, a 150 μm × 150 μm region was sampled from each TACS pattern in the MPM images, generating 142 collagen-related microscopic features via MATLAB 2016b (MathWorks, Natick, MA). The most informative features were selected using LASSO regression, and a weighted linear combination of these features produced the TCMF1 score [17].

TCMF2 scoring was performed on the digitized H&E images. From each TACS region, a 180 μm × 180 μm area was extracted to capture 179 nuclear features (Supplementary Methods, https://doi.org/10.1016/j.esmoop.2024.103990).

Following the same approach as for TCMF1, LASSO regression identified the most predictive nuclear features, which were then combined linearly with weights from the regression to compute the TCMF2 score.

Statistical analysis

Statistical analyses were conducted using IBM SPSS Statistics 25 (New York, NY) and R version 4.0.5 (R Foundation, Vienna, Austria) and. To identify the most informative features for TCMF1 and TCMF2, LASSO regression was applied. The TACS, TCMF1, and TCMF2 scores were integrated through a linear combination within a Coxn and univariate proportional hazards framework. Independent predictors were selected using multivariate Cox regression analyses with the likelihood ratio test. Prognostic accuracy was assessed via receiver operating characteristic (ROC) curve analysis, with areas under the curve (AUCs) calculated for evaluation. In the training cohort, the optimal cut-off value for risk classification was determined by maximizing the Youden index, which was then applied to categorize patients into high- and low risk groups across all cohorts. Survival associations were examined using Kaplan–Meier analysis alongside Cox regression models. All tests were two-sided, and statistical significance was defined as $P < 0.05$.

Ethics approval

Ethical clearance for this retrospective study was obtained from the Institutional Review Boards of Harbin Medical University Cancer Hospital (approval number: KY2020-11) and Fujian Medical University Union Hospital (approval number: 2020KJT010). All study procedures were conducted following the applicable regulatory standards and institutional guidelines.

Informed consent statement

Given the study's retrospective design, the requirement for written informed consent was exempted.

Results and Discussion

The comprehensive collagen signature serves as an independent prognostic indicator for survival in breast cancer patients.

Researchers developed the TACS score based on patterns TACS1 through TACS8, with the specific computation detailed in the Supplementary Methods (accessible at https://doi.org/10.1016/j.esmoop.2024.103990). Using automated computational tools, they then derived 142 microscopic collagen features associated with each TACS pattern and selected 14 key ones to form the TCMF1 score; its formula is also provided in the Supplementary Methods. Integrating early-stage collagen microscopic alterations (captured by TCMF1) with later macroscopic structural rearrangements (via TACS) enhanced prognostic accuracy, underscoring the importance of incorporating full collagen-related data for reliable outcome prediction [17].

Additionally, automated algorithms extracted 179 nuclear features from regions adjacent to TACS1-8, from which 17 stable features were chosen to create the TCMF2 score

The TCMF2 calculation is outlined in the Supplementary Methods. The full collagen signature was generated by integrating the TACS, TCMF1, and TCMF2 scores, with the combined formula (TACS + TCMF1 + TCMF2) available in the Supplementary Methods.

Univariate Cox regression showed that the full collagen signature score, alongside molecular subtype, tumor size, lymph node involvement, clinical stage, chemotherapy, and radiotherapy, significantly correlated ($P < 0.05$) with

disease-free survival (DFS) in the training group. In multivariate Cox analysis adjusting for these factors, the signature emerged as an independent predictor of DFS across all three cohorts Receiver operating characteristic (ROC) curve evaluation confirmed strong prognostic capability, yielding area under the curve (AUC) values of 0.897 in the training cohort, 0.864 in the internal validation cohort, and 0.849 in the external validation cohort **(Figure 1a)**.

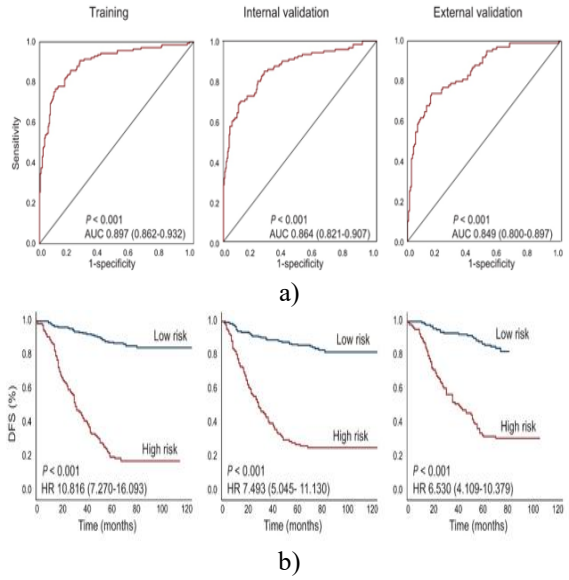

Figure 1. Assessment of the full collagen signature's ability to forecast patient outcomes. **(a)** ROC curves illustrating how well the complete collagen signature predicts disease-free survival (DFS) in the training, internal validation, and external validation groups. **(b)** Kaplan–Meier plots showing survival differences based on collagen signature scores across the three groups. Numbers in parentheses denote 95% confidence intervals. AUC = area under the curve; HR = hazard ratio.

An optimal threshold of 0.187, identified using the highest Youden index from the training group, stratified patients into low-risk and high-risk categories. In the training group, 63.1% were deemed low-risk (compared to 66.2% in internal validation and 56.7% in external validation), while 36.9% fell into the high-risk category (33.8% internal; 43.3% external).

High-risk individuals experienced markedly inferior DFS relative to their low-risk counterparts **(Figure 1b)**. Five-year DFS rates in the training group reached 87.5% for low-risk patients but only 19.8% for high-risk ones. In the internal validation group, these figures were 86.4% (low-risk) versus 27.4% (high-risk); in the external validation group, 88.1% (low-risk) versus 33.9% (high-risk).

Association between the comprehensive collagen signature and TNM staging

Subgroup evaluations were conducted in the training set, aligning patients with the AJCC TNM classification. The comprehensive collagen signature proved capable of further distinguishing prognosis among individuals sharing the same TNM stage

Among those identified as low-risk, patients with stage I disease displayed disease-free survival (DFS) rates on par with low-risk stage II cases (P = 0.188). Similarly, low-risk stage II patients showed DFS akin to low-risk stage III patients (P = 0.087). In contrast, low-risk cases across stages I to III consistently outperformed high-risk stage I patients in terms of DFS **(Figure 2)**.

For high-risk individuals, stage I patients had DFS outcomes similar to high-risk stage II patients (P = 0.622), but high-risk stage II patients demonstrated superior DFS relative to high-risk stage III patients (P = 0.012) **(Figure 2)**.

These observations imply that a low-risk signature may effectively reclassify (or "downstage") certain patients from stage II or III, whereas a high-risk signature could reclassify (or "upstage") stage I patients to reflect greater risk. Similar patterns were evident in the internal and external validation sets.

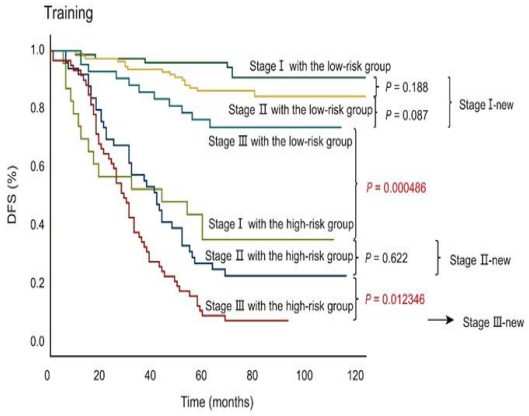

Figure 2. Influence of the comprehensive collagen signature on the traditional tumor–node–metastasis (TNM) staging framework in the training group.

P values were derived from log-rank tests, with significant results (P < 0.05) highlighted in red. DFS: disease-free survival.

Drawing on insights from the full collagen signature, researchers proposed a revised staging approach: low-risk cases from original stages I, II, and III were grouped together as "stage I-new"; high-risk cases from stages I and II were reassigned to "stage II-new"; and high-risk stage III cases were designated "stage III-new," reflecting their notably poorer DFS compared to the other categories.

In the standard TNM system, stage III patients exhibited significantly inferior DFS relative to stage I patients (referent stage I: HR 4.131, 95% CI 2.523–6.765, P = 1.72×10^{-8}), whereas no significant DFS difference existed between stage II and stage I patients (referent stage I: HR 1.642, 95% CI 0.991–2.718, P = 0.054; **(Table 1)**).

In contrast, the updated system integrating the collagen signature (I-new, II-new, III-new) achieved clearer prognostic separation, with progressively worsening DFS across the new stages (referent stage I-new: stage II-new HR 8.655, 95% CI 5.583–13.418, P = 4.99×10^{-22}; stage III-new HR 14.855, 95% CI 9.504–23.219, P = 2.37×10^{-32}; **Table 1**). This enhanced stratification ability was consistently confirmed in both internal and external validation groups.

Table 1. Prognostic value for disease-free survival (DFS) under the conventional TNM staging versus the revised staging system in the training group.

System	P value	HR (95% CI)
Modified staging system		
III with low-risk group +II with low-risk group + I-new (I with low-risk group		Reference
II-new (II with high-risk group + I with high-risk group)	4.99×10^{-22}	8.655 (5.583-13.418)
III-new (III with high-risk group)	2.37×10^{-32}	14.855 (9.504-23.219)
Current TNM staging system		
I		Reference
II	0.054	1.642 (0.991-2.718)
III	1.72×10^{-08}	4.131 (2.523-6.765)

P values were determined via log-rank tests, with significant differences highlighted in italics.

When the comprehensive collagen signature was integrated into the conventional TNM staging framework, patients in the new stage I-new group exhibited disease-free survival (DFS) similar to that of the original stage I group. However, those in stage II-new and stage III-new experienced notably poorer DFS compared to their counterparts in the original stage II and stage III categories, respectively.

In the training cohort, median DFS reached 78.0 months for stage I-new (77 months internal validation; 80 months external validation), versus 76.0 months for original stage I (78 months internal; 79.5 months external). For stage II-new, it was 42.0 months (40.5 months internal; 57 months external), compared to 75.0 months in original stage II (68 months internal; 78 months external). Stage III-new patients had a median DFS of 29.0 months (21 months internal; 20 months external), against 43.0 months for original stage III (41 months internal; 37 months external).

These shifts in DFS patterns highlight the revised system's superior risk stratification, particularly for patients originally in stages II and III.

Furthermore, the updated staging approach outperformed the standard TNM system in terms of area under the curve (AUC): 0.843 vs. 0.683 in the training cohort; 0.792 vs. 0.661 in internal validation; and 0.793 vs. 0.646 in external validation. This confirms the modified system's stronger capability for patient prognostication.

Distribution of TACS1–8 across the revised and traditional TNM staging frameworks

Researchers evaluated the average occurrence rates of each TACS type (TACS1–8) in both the conventional and revised TNM staging systems across all three cohorts. TACS1, TACS4, and TACS7—previously linked to favorable outcomes in earlier work[16]—along with TACS5, TACS6, and TACS8 (associated with unfavorable prognosis), displayed consistent shifts in frequency when moving from the standard TNM system to the new modified system.

In the stage I-new group, the prevalence of the favorable TACS1, TACS4, and TACS7 showed a modest upward trend in all cohorts compared to original stage I (with the exception of a slight decline in TACS7 in the internal validation cohort), whereas the unfavorable TACS5, TACS6, and TACS8 exhibited a mild downward trend relative to stage I. Conversely, in stage II-new and III-new groups, the favorable TACS types (1, 4, 7) decreased in frequency compared to their original stage II and III

counterparts, while the unfavorable types (5, 6, 8) increased. Notably, TACS4 (strongly tied to good prognosis) and TACS6 (strongly tied to poor prognosis) demonstrated the most pronounced changes during the transition from original stage II to II-new and from stage III to III-new, respectively[16].

Clinical utility of the revised staging system

Application of the standard TNM staging to the 941 patients categorized 253 as stage I (suitable for less intensive therapy), 438 as stage II, and 250 as stage III (requiring more intensive therapy). The revised system, however, reallocated 588 patients to stage I-new, 214 to stage II-new, and 139 to stage III-new.

This reclassification affected 48.4% (455/941) of patients overall: 6.4% (60/941) originally in stage I were upstaged to II-new, whereas 42.0% (395/941)—comprising 284 from original stage II and 111 from stage III—were downstaged to I-new.

Of the 60 upstaged from stage I to II-new, recurrence was accurately forecasted in 38 cases. Among the 395 downstaged from stages II/III to I-new, non-recurrence was correctly predicted in 327 patients.

Table 2 outlines tumor features according to reclassification status ('upstaged', 'no change', or 'downstaged') between the traditional TNM and the new system. The comprehensive collagen signature tended to upstage early breast cancers (tumors <2 cm, node-negative, diagnosed as stage I), while downstaging more advanced cases, including 305 with larger tumors (≥2 cm), 246 with positive nodes, and all 395 originally in stages II and III.

Figure 3 illustrates net-benefit comparisons, revealing that the revised staging system provided greater clinical net benefit than the conventional TNM approach.

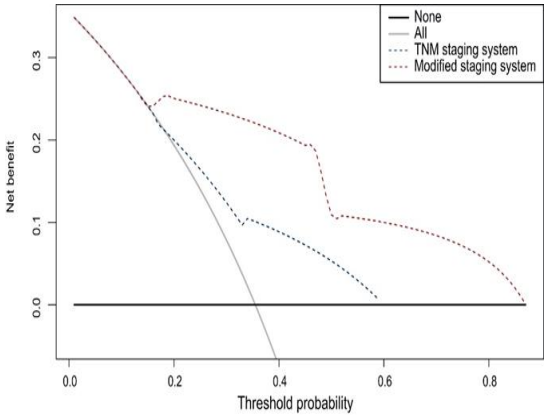

Figure 3. Decision curve analysis contrasting the traditional tumor–node–metastasis (TNM) staging approach with the newly proposed revised staging system.

Table 2. Characteristics of tumors based on reclassification ('upstaged', 'unchanged', or 'downstaged') when shifting from the conventional TNM framework to the updated staging system.[a]

Characteristics	Downstage (n = 395), n (%)	All patients (N = 941), n	Upstage (n = 60), n (%)	No change (n = 486), n (%)
Molecular subtype				
Luminal A	84 (41.4)	203	14 (6.9)	105 (51.7)
Luminal B	178 (44.2)	403	23 (5.7)	202 (50.1)
HER2-enriched	83 (44.6)	186	12 (6.5)	91 (48.9)
Triple-negative	50 (33.6)	149	11 (7.3)	88 (59.1)
Age (years)				
≤50	233 (44.1)	528	28 (5.3)	267 (50.6)
>50	162 (39.2)	413	32 (7.8)	219 (53.0)
Nodal status				
0	149 (32.0)	466	60 (12.9)	257 (55.1)
1-3	137 (59.6)	230	0 (0.0)	93 (40.4)
≥4	109 (44.5)	245	0 (0.0)	136 (55.5)
Tumor size (cm)				
≤2	90 (21.3)	423	60 (14.2)	273 (64.5)
2-5	281 (60.0)	468	0 (0.0)	187 (40.0)
>5	24 (48.0)	50	0 (0.0)	26 (52.0)

Histological grade				
G1	48 (39.0)	123	13 (10.6)	62 (50.4)
G2	239 (41.9)	571	40 (7.0)	292 (51.1)
G3	108 (43.7)	247	7 (2.8)	132 (53.5)
Clinical stage				
I	0 (0.0)	253	60 (23.7)	193 (76.3)
II	284 (64.8)	438	0 (0.0)	154 (35.2)
III	111 (44.4)	250	0 (0.0)	139 (55.6)
HER2 status				
Negative	252 (40.2)	627	40 (6.4)	335 (53.4)
Positive	143 (45.5)	314	20 (6.4)	151 (48.1)
Progesterone receptor status				
Negative	169 (40.8)	414	30 (7.3)	215 (51.9)
Positive	226 (42.9)	527	30 (5.7)	271 (51.4)
Estrogen receptor status				
Negative	133 (39.7)	335	23 (6.9)	179 (53.4)
Positive	262 (43.2)	606	37 (6.1)	307 (50.7)

TNM, tumor–node–metastasis; HER2, human epidermal growth factor receptor 2.
[a]% indicates the 'downstage', 'upstage', or 'no stage' distribution within the same subtype.

The tumor–node–metastasis (TNM) staging system, which is currently the most widely adopted framework for cancer classification, characterizes disease severity based on tumor size, lymph node involvement, and the presence of distant metastases [3]. Although this approach captures fundamental anatomical information, it fails to account for tumor biology and does not directly reflect the invasive potential of cancer cells. As a result, patients harboring small but biologically aggressive tumors may be insufficiently treated [21, 22].

Collagen, the predominant structural component of the extracellular matrix (ECM) within the tumor microenvironment, plays a critical role in regulating cancer invasion and progression by providing biomechanical cues [23–26]. In a seminal mouse study, Provenzano *et al*. [14] identified three tumor-associated collagen signatures (TACS1–3) surrounding tumors. Subsequent studies in human breast cancer confirmed that TACS3 was associated with reduced patient survival [15]. Building on these findings, we expanded the TACS classification in 2021 by identifying five additional collagen patterns (TACS4–8) using multiphoton microscopy (MPM), and demonstrated that the full TACS1–8 model exhibited strong prognostic value.[16] Notably, the corresponding collagen microscopic feature score (TCMF1) was later shown to be an independent prognostic factor [17]. Together, TACS and TCMF1 capture complementary macroscopic and microscopic collagen characteristics that reflect breast cancer progression [17].

Tumor cells actively remodel the ECM by inducing matrix metalloproteinase production while simultaneously promoting collagen deposition and fiber reorganization. Increased collagen accumulation and crosslinking during tumor progression enhance matrix stiffness [27], which in turn alters cancer cell behavior and promotes invasiveness [28]. The spatial orientation of collagen fibers relative to tumor cells critically influences migration: fibers aligned perpendicular to the tumor boundary facilitate directional movement and invasion[15], whereas fibers arranged parallel to the tumor boundary restrict migration and act as physical barriers[16].

Distinct TACS patterns reflect these biological mechanisms. TACS1 consists of curved collagen fibers encasing early tumor foci and can confine tumor growth during initial development [29]. TACS4 features reticular collagen bundles aligned parallel to the tumor edge, effectively limiting outward expansion [16, 30]. In contrast, TACS5 is characterized by straightened, vertically aligned fibers that form permissive tracks for tumor invasion [15, 16, 31, 32], while TACS6 displays

disorganized collagen alignment that enables multidirectional tumor cell movement in the absence of a defined boundary [33]. TACS7 comprises densely packed collagen fibers at the invasive front, forming barriers that restrain invasion [16], whereas TACS8 exhibits sparse collagen distribution associated with ECM degradation at the invasion front, thereby facilitating early tumor spread [34]. Collectively, these patterns demonstrate that the spatial relationship between tumor cells and collagen architecture governs migratory behavior and clinical outcomes. Accordingly, TACS1, TACS4, and TACS7 are associated with favorable prognosis due to their inhibitory effects on migration, whereas TACS5, TACS6, and TACS8 are linked to poor prognosis by promoting invasion.

In the present study, only modest changes in the distribution of favorable (TACS1,4,7) and unfavorable (TACS5,6,8) patterns were observed when reclassifying patients from stage I to stage I-new, which may explain why disease-free survival (DFS) remained comparable between these groups. In contrast, patients assigned to stage II-new and III-new demonstrated a marked increase in poor prognostic TACS patterns alongside a reduction in favorable patterns—particularly notable shifts in TACS4 and TACS6—which may underlie their worse DFS compared with patients staged as II or III using the conventional TNM system. TACS2 and TACS3 showed weaker prognostic associations in previous studies [16], and their effects may therefore be masked by the stronger influence of TACS4–8 during stage reclassification.

Incorporating collagen-related information offers several advantages. First, it compensates for the lack of biological context in the traditional TNM system, which may inadequately represent tumor aggressiveness. Notably, some small tumors exhibit highly invasive behavior, indicating that biological features may outweigh anatomical extent in predicting outcomes [21, 35]. Consistent with this concept, inclusion of the complete collagen signature led to the reclassification and “upstaging” of 60 patients with anatomically small tumors, of whom 38 subsequently experienced recurrence. These findings highlight the clinical utility of collagen-derived biological information. The modified staging system therefore outperformed the conventional TNM framework in prognostic discrimination, underscoring the central role of collagen remodeling in tumor invasion and its relevance for treatment decision-making. Tumors exhibiting organized collagen architecture may have limited invasive capacity and could potentially be managed with less aggressive therapy, reducing treatment-related morbidity [30]. Conversely, tumors characterized by dense and disordered collagen matrices may require intensified treatment strategies to mitigate the risk of recurrence or metastasis [29, 36].

Second, given the ubiquitous role of collagen in the ECM across multiple organ systems, collagen-based biomarkers may have broader applicability for risk stratification in other cancer types, offering new avenues for translational and mechanistic research. Third, the use of formalin-fixed paraffin-embedded specimens allows for convenient transport from remote collection sites to centers equipped with MPM, making this collagen-focused histological approach cost-effective and feasible for implementation in resource-limited settings. Continued development of miniaturized MPM devices may further support routine clinical application, even in well-resourced healthcare systems.

In this study, the complete collagen signature was derived through colocalized analysis of MPM and H&E images. However, advances in virtual staining based on generative adversarial networks offer the potential to convert MPM images into virtual H&E images using unsupervised deep learning, enabling straightforward extraction of collagen signatures directly from virtually stained slides. Encouraging evidence supports this approach: Picon *et al*. [37] reported superior diagnostic performance of virtual H&E images derived from MPM in human colon lesions compared with conventional H&E analysis. Additionally, blinded diagnostic evaluations of MPM, virtual-stained, and standard H&E images in cerebral cavernous malformations demonstrated comparable accuracy between virtual-stained and traditional H&E images, confirming the reliability and clarity of virtual staining techniques [38].

Conclusion

The complete collagen signature serves as an independent prognostic marker in breast cancer. When incorporated into the TNM staging framework, it captures tumor biological behavior and substantially improves stratification of patient outcomes.

Acknowledgments: We thank all members of Chen Lab for their suggestions and critical feedback.

Conflict of Interest: None

Financial Support: This research was funded by the National Natural Science Foundation of China [grant numbers 82171991 and 81700576] and the Natural Science Foundation of Fujian Province [grant numbers 2024J02013, 2023J01504, 2023J011125, 2022J01170, and 2024J01624].

Ethics Statement: None

References

1. Sung, H. Ferlay, J. Siegel, R.L. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021; 71:209-249.
2. Singletary, S.E. · Allred, C. Ashley, P. Staging system for breast cancer: revisions for the 6th edition of the AJCC Cancer Staging Manual. Surg Clin North Am. 2003; 83:803-819.
3. Weiss, A. · King, T.A. Hunt, K.K. Incorporating biologic factors into the American Joint Committee on Cancer breast cancer staging system: review of the supporting evidence. Surg Clin North Am. 2018; 98:687-702.
4. Cserni, G. · Chmielik, E. · Cserni, B. The new TNM-based staging of breast cancer. Virchows Arch. 2018; 472:697-703.
5. He, J. · Tsang, J.Y. · Xu, X. AJCC 8th edition prognostic staging provides no better discriminatory ability in prognosis than anatomical staging in triple negative breast cancerBMC Cancer. 2020; 20:18.
6. Weiss, A. · Chavez-MacGregor, M. · Lichtensztajn, D.Y. Validation study of the American Joint Committee on Cancer eighth edition prognostic stage compared with the anatomic stage in breast cancer. JAMA Oncol. 2018; 4:203-209.
7. Jiang, W. · Yu, X. · Dong, X. A nomogram based on collagen signature for predicting the immunoscore in colorectal cancer. Front Immunol. 2023; 14, 1269700.
8. Conklin, M.W. · Gangnon, R.E. · Sprague, B.L. Collagen alignment as a predictor of recurrence after ductal carcinoma in situ. Cancer Epidemiol Biomarkers Prev. 2018; 27:138-145.
9. Xu, L. Si, H. Su, H. The number of metastatic lymph nodes is more predictive of prognosis than location-based N stage for nonsmall cell lung cancer: a retrospective cohort study. Int J Surg. 2023; 109:4126-4134.
10. Li, H. · Lin, D. · Yu, Z. A nomogram model based on the number of examined lymph nodes-delated signature to predict prognosis and guide clinical therapy in gastric cancer. Front Immunol. 2022; 13, 947802.
11. Tan, Z. Chen, X. · Zuo, J. Comprehensive analysis of ScRNA-seq and bulk RNA-seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model. J Transl Med. 2023; 21:223.
12. Chen, J. Li, Z. Han, Z. Prognostic value of tumor necrosis based on the evaluation of frequency in invasive breast cancer. BMC Cancer. 2023; 23:530.
13. Kistenev, Y.V. · Vrazhnov, D.A. · Nikolaev, V.V. Analysis of collagen spatial structure using multiphoton microscopy and machine learning methods. Biochemistry (Mosc). 2019; 84:S108-S123.
14. Provenzano, P.P. · Eliceiri, K.W. · Campbell, J.M. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 2006; 4:38.
15. Conklin, M.W. · Eickhoff, J.C. · Riching, K.M. Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am J Pathol. 2011; 178:1221-1232.
16. Xi, G. · Guo, W. · Kang, D. Large-scale tumor-associated collagen signatures identify high-risk breast cancer patients. Theranostics. 2021; 11:3229-3243.
17. Xi, G. · Qiu, L. · Xu, S. Computer-assisted quantification of tumor-associated collagen signatures to improve the prognosis prediction of breast cancer. BMC Med. 2021; 19:273.
18. Nafe, R. · Franz, K. · Schlote, W. Morphology of tumor cell nuclei is significantly related with survival time of patients with glioblastomas. Clin Cancer Res. 2005; 11:2141-2148.
19. Wang, X. · Janowczyk, A. · Zhou, Y. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. Sci Rep. 2017; 7, 13543
20. Riedl, P. · Sapudom, J. · Clemens, C. Phenotype switching of breast cancer cells upon matrix interface crossing. ACS Appl Mater Interfaces. 2023; 15:24059-24070.
21. Wo, J.Y. · Chen, K. · Neville, B.A. Effect of very small tumor size on cancer-specific mortality in

node-positive breast cancer. J Clin Oncol. 2011; 29:2619-2627.
22. Comen, E.A. · Norton, L. · Massagué, J. Breast cancer tumor size, nodal status, and prognosis: biology trumps anatomy. J Clin Oncol. 2011; 29:2610-2612.
23. Nelson, M.S. · Liu, Y. · Wilson, H.M. Multiscale label-free imaging of fibrillar collagen in the tumor microenvironment. Methods Mol Biol. 2023; 2614:187-235.
24. Papanicolaou, M. · Parker, A.L. · Yam, M. Temporal profiling of the breast tumour microenvironment reveals collagen XII as a driver of metastasis. Nat Commun. 2022; 13:4587.
25. He, X. · Lee, B. · Jiang, Y. Cell-ECM interactions in tumor invasion. Adv Exp Med Biol. 2016; 936:73-91.
26. Sapudom, J. · Pompe, T. Biomimetic tumor microenvironments based on collagen matrices. Biomater Sci. 2018; 6:2009-2024.
27. Bera, K. · Kiepas, A. · Zhang, Y. The interplay between physical cues and mechanosensitive ion channels in cancer metastasis. Front Cell Dev Biol. 2022; 10, 954099.
28. Nia, H.T. · Munn, L.L. · Jain, R.K. Physical traits of cancer. Science. 2020; 370, eaaz0868.
29. Poonja, S. Forero Pinto, A. Lloyd, M.C. Dynamics of fibril collagen remodeling by tumor cells: a model of tumor-sssociated collagen signatures. Cells. 2023; 12:2688.
30. Tian, C. Huang, Y. Clauser, K.R. Suppression of pancreatic ductal adenocarcinoma growth and metastasis by fibrillar collagens produced selectively by tumor cells. Nat Commun. 2021; 12:2328.
31. Nelson, M.T. · Short, A. · Cole, S.L. Preferential, enhanced breast cancer cell migration on biomimetic electrospun nanofiber 'cell highways'. BMC Cancer. 2014; 14:825.
32. Han, W. · Chen, S. · Yuan, W. Oriented collagen fibers direct tumor cell intravasation. Proc Natl Acad Sci U S A. 2016; 113:11208-11213.
33. Chen, W. · Gu, T. · Chen, Q. Extracellular matrix remodelling and stiffening contributes to tumorigenesis of salivary carcinoma ex pleomorphic adenoma—a study based on patient-derived organoids. Cell Biosci. 2023; 13:122
34. Cai, R. · Tressler, C.M. · Cheng, M. Primary breast tumor induced extracellular matrix remodeling in premetastatic lungs. Sci Rep. 2023; 13, 18566.
35. Hernandez-Aya, L.F. · Chavez-Macgregor, M. · Lei, X. Nodal status and clinical outcomes in a large cohort of patients with triple-negative breast cancer. J Clin Oncol. 2011; 29:2628-2634.
36. Deng, J. · Kang, Y. · Cheng, C.C. DDR1-induced neutrophil extracellular traps drive pancreatic cancer metastasis. JCI Insight. 2021; 6, e146133
37. Picon, A. · Terradillos, E. · Sánchez-Peralta, L.F. Novel pixelwise co-registered hematoxylin-eosin and multiphoton microscopy image dataset for human colon lesion diagnosis. J Pathol Inform. 2022; 13, 100012.
38. Wang, S. Li, Y. Xu, Y. Resection-inspired histopathological diagnosis of cerebral cavernous malformations using quantitative multiphoton microscopy. Theranostics. 2022; 12:6595-6610.