Breast Cancer Research

REVIEW

Open Access

The association between tumor-stromal collagen features and the clinical outcomes of patients with breast cancer: a systematic review

Samane Heydari[1,2], Fatemeh Tajik[1], Sadegh Safaei[1,2], Fereshteh Kamani[3], Babak Karami[4], Shima Dorafshan[1,2], Zahra Madjd[1,2*] and Roya Ghods[1,2*]

Abstract

Background The tumor microenvironment (TME), particularly the extracellular matrix (ECM), plays a crucial role in regulating breast cancer progression. Among ECM components, collagen type I—accounting for over 90% of fibrillar collagen in the human body—is the primary structural component of the tumor ECM. It critically modulates tumor cell behavior, influencing migration, invasion, and therapy resistance. The structural organization of collagen type I fibers can significantly impact clinical outcomes.

Methods This systematic review aimed to assess the association between tumor-stromal collagen type I characteristics and clinical outcomes in breast cancer. A comprehensive search strategy identified studies from major databases, which were appraised using quality assessment tools. Data on collagen quantity, morphology, alignment, and organization were extracted and analyzed to explore their relationship with survival, metastasis, therapy resistance, and clinical characteristics of breast cancer.

Results Our analysis revealed that increased collagen density—particularly with an organized/aligned fiber orientation—was strongly associated with poor prognosis. Specifically, increased intratumoral collagen quantity was linked to reduced overall survival (HR = 7.84, $p = 0.031$). Stage III tumors exhibiting elevated collagen uniformity showed higher metastasis rates ($p = 0.004$), and HER2$^+$ tumors with high collagen content correlated with resistance to HER2-targeted therapies ($p < 0.05$). Furthermore, higher collagen curviness was associated with better outcomes, including a reduced recurrence risk (HR = 0.77, $p < 0.001$). Subtype-specific trends emerged as ER/PR-negative tumors more frequently exhibited a perpendicular collagen arrangement ($p = 0.02$), whereas ER/PR-positive tumors showed elevated COL1A1 expression ($p < 0.0001$). Despite these patterns, the heterogeneity of study methodologies and the complexity of the tumor microenvironment highlight the need for unified frameworks to advance clinical translation.

*Correspondence:
Zahra Madjd
majdjabari.z@iums.ac.ir; zahra.madjd@yahoo.com
Roya Ghods
ghods.ro@iums.ac.ir; rghods77@yahoo.com
Full list of author information is available at the end of the article

Conclusions This review highlights the prognostic significance of tumor-stromal collagen characteristics in breast cancer, suggesting that future research should focus on the molecular mechanisms underlying collagen remodeling and its potential as a cancer biomarker and therapeutic target.

Keywords Collagen type I, Breast cancer, Extracellular matrix, Tumor microenvironment, Prognostic marker, Systematic review

Introduction

Breast cancer, a significant global health challenge, affects 2.1 million women every year and results in the death of 30% of those diagnosed [1]. Increasing evidence suggests that the tumor microenvironment (TME) plays a fundamental role in tumor progression across various cancer types, including breast cancer [2]. Identifying the specific contributions of each TME component could lead to more effective cancer management strategies [3]. One hallmark feature of the TME in desmoplastic cancers, such as breast cancer, is the extracellular matrix (ECM) [2]. The ECM in breast tissue is a dynamic network of structural proteins, including collagens, elastin, periostin, laminins, fibronectin, osteonectin, thrombospondin, osteopontin, entactin, tenascin, and glycosaminoglycans, and among these, collagens are the most abundant protein in ECM [4], and play a crucial role in cancer cell survival and metastasis [5].

More than 28 types of collagens have been identified in humans and are categorized by structure and function. Fibrillar collagens, such as types I, II, and III, provide tensile strength, while collagens associated with fibrils, like types IX and XII, regulate fibril assembly. Network-forming collagens, such as type IV, constitute basement membranes, and other collagens, including types XIII and XVII, anchor cells to the ECM [4]. Collagen type I, the focus of this review, constitutes 90% of the fibrillar collagen in breast cancer stroma, making it the dominant component observed and a key determinant of matrix biophysical properties [6].

Collagen type I, referred to as collagen in this study, can affect tumor cells behavior via its biochemical properties [7]. Furthermore, the ability of cancer cells to sense and respond to mechanical signals has attracted increasing attention in recent years [8]. During cancer initiation and progression, there is a continuous reciprocal interplay between cancer cells and collagen, both within and around the tumor. This dynamic interplay results in changes in various collagen fiber parameters [9]. A variety of techniques are used to evaluate collagen parameters in breast cancer.

Methodologies for investigating collagen fiber features in breast cancer research

In addition to traditional methods such as real-time PCR (which measures mRNA expression of collagen type I, alpha 1, *COL1A1*) and immunohistochemistry (IHC, which detects *COL1A1* protein) [10, 11], collagen structural features (such as quantity and organization) can also be assessed through different staining and imaging techniques, as well as computational analyses.

Imaging techniques

Second Harmonic Generation (SHG) microscopy SHG microscopy is a powerful label-free optical technique widely employed for the specific imaging of collagen fibers. This method harnesses the unique structure of fibrillar collagens to generate a high-contrast signal without the need for exogenous stains or labels. The fundamental principle of SHG involves the interaction of two photons with the same frequency from an incident near-infrared laser beam with the non-centrosymmetric arrangement of collagen molecules, resulting in the emission of a single photon with exactly double the frequency (half the wavelength) of the excitation light. This intrinsic property allows for highly specific visualization of collagen [12, 13].

SHG microscopy offers several key advantages for the study of collagen in biological systems. Its ability to resolve submicron features enables the detailed visualization of collagen fiber architecture and organization [14]. The technique is versatile and can be applied to fresh, fixed, or formalin-fixed paraffin-embedded (FFPE) tissue samples, making it adaptable to various research and potential clinical applications [13]. Moreover, Polarimetric SHG (P-SHG) enhances quantitative analysis by resolving collagen fiber orientation and molecular alignment, critical for studying structural changes in pathologies like breast cancer [13, 15].

Polarized Light Microscopy (PLM) PLM is a technique that capitalizes on the intrinsic birefringence of collagen fibers to visualize their organization within tissues. Birefringence is an optical property exhibited by anisotropic

materials where the refractive index of the material varies depending on the polarization and propagation direction of light passing through it. Collagen fibers, due to their highly ordered molecular arrangement, possess this characteristic. PLM, when combined with Sirius Red staining, offers a cost-effective 2D visualization of collagen birefringence, suitable for broad screening. However, unlike SHG (which provides 3D label-free imaging), PLM lacks depth resolution and molecular specificity. However, this accessibility makes PLM a valuable tool, particularly in resource-limited settings or for initial broad screening of tissue samples [16].

Quantitative Phase Imaging (QPI) QPI maps the optical path length difference (OPD) in tissue samples, providing insights into their nanoarchitecture without the need for stains. It works by measuring light phase shifts using interferometry, offering label-free imaging that avoids staining-related artifacts. Color Spatial Light Interference Microscopy (cSLIM) is an enhancement of QPI that allows the analysis of stained tissues. cSLIM integrates with stained samples by computationally separating stain-induced phase shifts from intrinsic tissue properties, enabling simultaneous traditional and label-free analysis. Therefore, it integrates seamlessly with current pathology workflows, enabling both traditional stain-based analysis and advanced stain-independent assessments via mathematical corrections. While QPI (and cSLIM) detects refractive index variations in dense ECM structures (including collagen), it lacks SHG's specificity. Collagen is inferred based on structural morphology. With compatibility for stained tissues and high throughput potential, cSLIM shows great promise for clinical and pathology applications, bridging traditional and innovative methodologies effectively [17, 18].

Staining techniques

Hematoxylin and Eosin (H&E) Staining H&E staining enables the visual identification of collagen-rich stroma in tissue samples due to their characteristic eosinophilic (pink/red) staining. This routine and cost-effective method allows for an initial assessment of the collagenous stroma. Crucially, H&E-stained slides can undergo computational analysis to yield quantitative data on collagen. While techniques like SHG offer greater specificity, the widespread availability of H&E makes computational approaches valuable for leveraging existing resources in collagen studies [17, 19].

Masson's trichrome staining Masson's trichrome staining is a histological technique utilized to assess the composition of collagen by specifically highlighting collagen fibers in blue, allowing for its clear visual identification within the tissue. This method is crucial as it enables the measurement of collagen area percentage in tissue sections through subsequent image analysis. While providing valuable quantitative data on bulk collagen content for studying various pathological conditions like breast cancer, it primarily reveals the overall amount rather than detailed fibrillar-level structural organization, for which techniques like SHG are more suited [20, 21].

Sirius red and picrosirius red staining Sirius red staining and the closely related picrosirius red staining are collagen-specific histochemical techniques that function by intensely binding to collagen fibers and enhancing their natural birefringence when viewed under polarized light, making them prominent methods for collagen structural analysis. They can be used to visualize collagen and are frequently combined with polarization microscopy to offer detailed insights into the organization of the collagen network within tissue samples. Picrosirius red staining, a variant of Sirius red, is prepared by mixing Sirius Red with a saturated aqueous solution of picric acid. When viewed under light microscopy, picrosirius red stains collagen fibers red, while epithelial cells appear yellow. This staining method under polarized light intensifies the birefringence of collagen, allowing for a more detailed assessment of its structure [22, 23].

Computational analysis

Following the acquisition of images through various staining and microscopy techniques, a range of computational methods are essential for extracting quantitative information about collagen fiber features. Automated tools like CT-FIRE (for fiber extraction and morphometry) and CurveAlign (for spatial alignment analysis) are critical for processing SHG microscopy images, enabling precise quantification of fiber length, width, alignment, and orientation relative to tumor-stroma interfaces [24]. Beyond these automated approaches, more sophisticated analytical techniques like curvelet transform are also employed. The curvelet transform is a sophisticated mathematical tool that is particularly well-adapted for analyzing images containing curve-like features, such as the fibrillar structure of collagen captured in SHG microscopy images [25]. By applying the curvelet transform to SHG images, researchers can then derive several important quantitative features of individual collagen fibers such as orientation, length, and curvature, providing a detailed characterization of the collagen network [26].

Other computational approaches provide valuable insights into collagen organization. Texture analysis,

which includes both first-order (evaluating the properties of individual pixels within an image) and second-order (assessing the relationships between pairs or groups of pixels) statistical techniques, enables the extraction of features indicative of structural changes in the collagen network. A key tool within this domain is the Gray Level Co-occurrence Matrix (GLCM), a statistical approach that examines the spatial relationships between pixel pairs to analyze the texture of collagen fibers. Parameters derived from GLCM, such as correlation and energy, provide significant distinctions in collagen fiber patterns. GLCM quantifies collagen texture via spatial pixel relationships. For example, GLCM energy reflects uniformity, while correlation indicates linear patterns. Coherence, a measure of fiber alignment, is calculated separately using structure tensor analysis [25].

Another valuable computational approach for quantitatively evaluating the overall orientation of collagen fibers within SHG microscopy images is the Fast Fourier Transform (FFT) analysis. SHG image of collagen fibers is composed of various directional elements, FFT decomposes SHG images into spatial frequencies. Dominant fiber alignment produces a frequency spectrum with intensity peaks perpendicular to the fiber direction. An ellipse fitted to the spectrum quantifies alignment: a high aspect ratio (elongated ellipse) indicates strong alignment, while a circular shape reflects randomness. The aspect ratio defines the collagen orientation index [27, 28]. To derive a single, quantifiable measure of this overall orientation, an ellipse is often fitted to the distribution of frequencies in the spectrum. The shape of this fitted ellipse directly reflects the degree of fiber alignment: a highly elongated ellipse indicates a strong preference for a particular orientation, whereas a more circular ellipse suggests a lack of dominant orientation. From the parameters of this ellipse, a collagen orientation index is calculated, providing a numerical value that characterizes the overall alignment of collagen fibers within the image. This allows for an objective and comparative assessment of collagen architecture in different biological contexts [29].

The integration of computational power extends to more advanced analytical paradigms. Machine learning and supervised learning algorithms are increasingly used to quantify and score the alignment of collagen fibers concerning adjacent tumor-stromal boundaries, to predict patient survival outcomes [30]. It is worth noting that despite the advancements in automated and complex computational methods, manual quantification of collagen fiber angles using tools like ImageJ remains a relevant approach in certain research studies [31, 32].

In addition to the diverse techniques employed, cancer-related modifications of collagen have been assigned various terms in different studies. In this review, the authors classify these terms into five distinct categories. The first category is alteration of the quantity or density of collagen in the TME, which is referred to by various labels, including the area of collagen I, density, expression, deposition, content, and intensity. Collagen quantity is influenced by both biosynthesis and biodegradation. Moreover, collagen density is influenced by covalent intermolecular cross-links within collagen fibers [4]. The second category of collagen that could be affected during cancer initiation and progression is indicated by fiber morphology, encompassing fiber width, length, and straightness/curviness [24]. The third feature is the intrinsic microstructural properties of individual collagen fibers, including fibril diameter, spacing, and order/disorder packing, which can be determined by the ratio of forward to backward (F/B) propagating SHG signals [12]. When a laser beam used for SHG microscopy interacts with the collagen fibers of the studied tissue sample, it generates SHG signals that can be emitted either in the same direction as the incident light (forward) or in the opposite direction (backward). The F/B SHG ratio quantifies the relative intensity of these forward and backward SHG signals and reflects the diameter or packing arrangement of the collagen fibrils within the tissue [12]. Fibrils with thinner diameters or sparse packing structures may exhibit weaker forward SHG signals, resulting in a lower F/B ratio, which could be related to abnormal, dysregulated, and rapid synthesis of collagen fibers (Fig. 1) [33]. Notably, the F/B SHG ratio reflects the aforementioned fibril properties, not the diameter or arrangement of the fibers [12].

However, the arrangement of collagen fibers can change during cancer progression, and terms such as organization, orientation, alignment, directionality variation, anisotropy, and dispersion can be attributed to this variable. Therefore, the fourth category is dedicated to collagen fiber organization. Additionally, during breast cancer progression, the angle of stromal collagen fibers to the tumor boundary tends to become perpendicular, representing the fifth category [34]. Moreover, several studies have attempted to design prognostic indices based on various collagen features [25, 26, 35–37]. For example, researchers have defined different patterns of TME collagen I, known as tumor-associated collagen signatures (TACSs), during tumor evolution. For example, TACS-1 stands for dense collagen regions wrapped around emergent tumor foci, TACS-2 represents straightened collagen fibers stretched around a tumor, suggesting growth and increased volume, and TACS-3 is related to radially aligned collagen fibers associated with tumor cell invasion, potentially indicating invasive and metastatic growth (Fig. 2) [38].

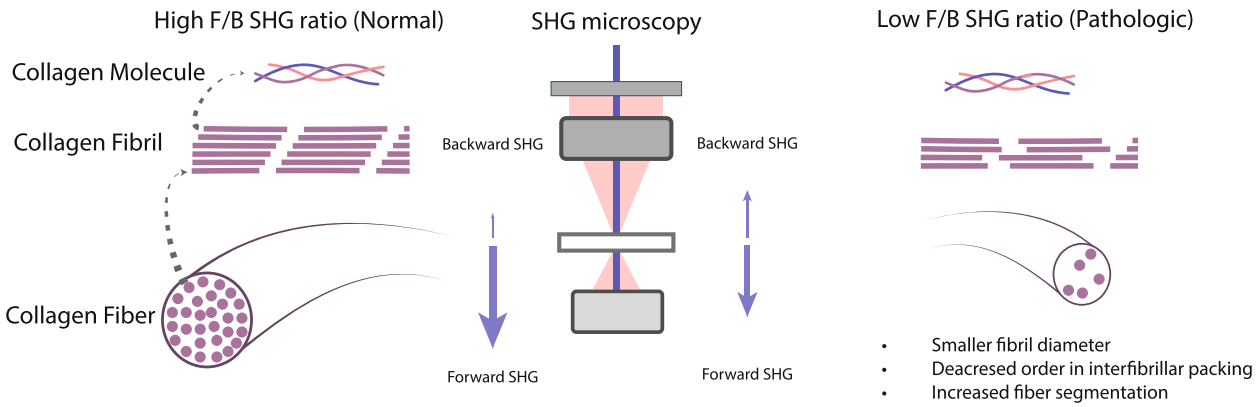

Fig. 1 Schematic representation of the relationship between collagen structure and the forward-to-backward (F/B) Second Harmonic Generation (SHG) ratio. This figure illustrates how the F/B SHG ratio, obtained through SHG microscopy, reflects the microstructural properties of collagen fibers. On the left, a high F/B SHG ratio, indicative of normal tissue, is associated with well-organized collagen fibrils forming thick, intact fibers. In contrast, on the right, a low F/B SHG ratio, characteristic of pathologic tissue, suggests smaller fibril diameters, decreased order in interfibrillar packing, and increased fiber segmentation. The center panel depicts the SHG microscopy process, where a laser beam interacts with collagen, generating forward and backward SHG signals. The ratio of these signals provides insights into the collagen's microstructural organization, which is altered in disease states

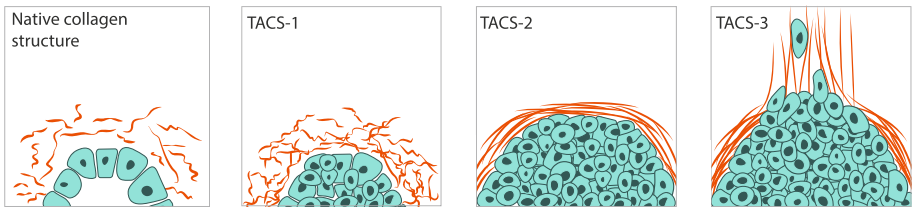

Fig. 2 Schematic depiction of Tumor-Associated Collagen Signatures (TACSs) during breast cancer progression. This figure illustrates the evolving patterns of collagen organization, known as TACSs, in breast cancer's tumor microenvironment. The "Native collagen structure" panel shows the baseline arrangement of collagen fibers. TACS-1 depicts dense collagen surrounding emerging tumor, TACS-2 illustrates straightened collagen fibers stretched around the tumor, suggesting tumor growth and increased volume. Finally, TACS-3 shows radially aligned collagen fibers associated with tumor cell invasion, potentially indicating invasive growth

Another study attempted to identify and introduce more TACSs at the invasive front of breast tumors. TACS-4 involves collagen fibers forming a network near expanding tumors, creating a distinct tumor boundary. TACS-5 pertains to collagen fibers that align directionally, facilitating one-way tumor cell migration without a clear boundary. TACS-6 indicates collagen fibers align chaotically, enabling multidirectional tumor cell migration without a clear boundary. TACS-7 refers to densely packed collagen fibers at the tumor invasion front that are mostly free of tumor cells, whereas TACS-8 refers to sparsely packed collagen fibers at the tumor invasion front that are mostly free of tumor cells, given that a single tumor may contain a variety of TACSs, a prognostic index has been formulated to reflect the composite TACS profile for an individual patient [35]. Some of the important collagen features analyzed in breast cancer research were summarized in Table 1.

Several studies have explored the associations of various indices or features of collagen with different aspects of breast cancer, such as hormone receptor (PR, ER, HER2) expression, tumor stage, grade, and the likelihood of recurrence, metastasis, therapy response, and survival. Some studies have attempted to develop a prognostic index based on various TME collagen features. Our objective was to conduct a systematic review that thoroughly reviewed these studies and uncovered intricate relationships between structural collagen features and breast cancer characteristics. Our analysis offers comprehensive insight into this complex association.

Materials and methods

Eligibility criteria

The eligibility criteria and research question have been established via the PICOs framework: "P" as population, "I" as intervention, "C" as comparator/control, "O"

Table 1 Collagen features in breast cancer research

Feature Category	Features	Definition	Measurement Methodologies
Quantity/Density	Total collagen density/volume	Overall collagen content in tissue, reflecting synthesis, deposition, or degradation.	SHG microscopy (collagen fiber volume, pixel ratio), Masson's trichrome/Sirius Red staining (collagen area percentage), IHC/real-time PCR (COL1A1 expression).
	Collagen intensity	Signal strength of collagen in stained or imaged samples.	Sirius Red staining (light microscopy), SHG microscopy (signal intensity).
Collagen Morphology	Fiber curvature/straightness	Degree of bending in fibers (curvature) or linearity (straightness).	SHG microscopy + CT-FIRE/CurveAlign (fiber tracing), ANN-based segmentation (length/width ratios).
	Fiber width/thickness	Diameter or thickness of individual collagen fibers.	SHG/PLM imaging + CT-FIRE, CurveAlign.
	Fiber length	End-to-end or contour length of collagen fibers.	SHG microscopy + CT-FIRE, CurveAlign, ANN segmentation.
Collagen Organization	Fiber alignment/anisotropy	Directional uniformity of fibers (0 = random, 1 = aligned).	SHG microscopy + CurveAlign (anisotropy coefficients), PLM (birefringence analysis).
	Collagen orientation index	Quantifies alignment (0 = random, 1 = linear).	FFT analysis + ellipse fitting (aspect ratio).
	GLCM parameters	Spatial relationships between pixels (e.g., correlation, energy).	Texture analysis of SHG/Sirius Red images (GLCM matrices).
	Ratio parameter (PSHG)	Proportion of fibrils conforming to polarized SHG model.	Polarization-dependent SHG (PSHG) + FFT analysis.
Collagen Microstructure	F/B SHG ratio	Forward-to-backward SHG signal ratio; reflects fibril diameter/packing.	SHG microscopy (emission direction analysis).
Perpendicular Relative to Tumor Boundary	TACS-3 signature	Straightened, aligned collagen fibers perpendicular to tumor boundary.	SHG microscopy + computational scoring (fiber angle, alignment).
Prognostic Indices	TACS scores (1–3, 1–8)	Composite scores of collagen signatures (e.g., TACS-3 = perpendicular fibers).	SHG imaging + ridge/LASSO regression (weighted collagen patterns).
	TCMF score	Microscopic collagen features (e.g., fiber length, texture).	Multiphoton microscopy + LASSO regression.
	IGNN score	Spatial heterogeneity of TACS patterns via graph neural networks.	SHG imaging + graph neural networks (spatial biomarker interactions).
	Collagen Prognostic Index	Combines directionality, density, and dispersion scores.	Picrosirius Red staining + scoring system.

Abbreviations: COL1A1 Collagen Type I Alpha 1, *F/B* Forward-to-Backward ratio, *FFT* Fast Fourier Transform, *GLCM* Gray-level co-occurrence matrix, *IHC* Immunohistochemistry, *IGNN* Intratumor graph neural network, *LASSO* least absolute shrinkage and selection operator, *PSHG* Polarization-dependent Second Harmonic Generation, *TACSs* Tumor-associated collagen signatures, *TCMFs* TACS corresponding microscopic features, *SHG* Second Harmonic Generation, *PLM* Polarized Light Microscopy

as outcome, and "S" as study type [39]. The following inclusion criteria were based on the PICO structures:

1) Observational studies (cross-sectional, case-control, and cohort studies) published in English after 1990 with available full-text and conference abstracts that evaluated the tumor-stromal collagen features of human breast cancer were included.
2) Studies with human breast cancer.
3) Studies have assessed the associations between TME collagen features of primary tumors in human breast cancer and the prognostic value of overall survival (OS), disease-free survival (DFS), relapse-free survival (RFS), and disease-specific survival (DSS)/cancer-specific survival (CSS).
4) Studies have assessed the associations between TME collagen features of primary tumors in human breast cancer and distant metastasis, recurrence, therapy resistance, tumor stage, tumor grade, and hormonal receptors of breast cancer cells (ER, PR, and HER2).
5) TME Collagen features include the following:
   - Collagen quantity: area, density, expression, deposition, content, and intensity of collagen fibers.
   - Collagen morphology: collagen fiber width, length, and straight/curviness.
   - Collagen organization: orientation, alignment, directionality variation, dispersion, and uniformity of collagen fibers.

- Angle between the fiber direction and the tumor boundary
- The microstructural properties of collagen are determined by the forward-SHG to backward-SHG ratio.
- Prognostic indices were developed on the basis of collagen features.

Investigations that met the following criteria were excluded:

1) Reviews, commentaries, case reports, and case series.
2) In vitro and in vivo studies
3) Studies not related to the topic of interest
4) Studies for which the full text was either unavailable or not in English.

Information sources

The search was conducted in four main electronic databases: PubMed/MEDLINE, Web of Science (WOS), Scopus, and Embase via www.Embase.com until June 1, 2023, and was updated on September 1, 2024. Furthermore, the search was extended to gray literature, including conference papers. We also reviewed the references within the selected studies to perform a comprehensive search and prevent any data from being overlooked.

Search strategy

The current study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [40]. The primary search terms employed in the search strategy were neoplasm, breast cancer, and collagen I. The search syntax was altered in other databases (Table S1). They were limited to English, which was published after 1990.

Selection process

Three steps were performed to include a paper in the study. Initially, duplicates were eliminated using EndNote software (version X9.3.3; Thomson Reuters, Philadelphia, U.S.A.) and manual searching. Two independent authors (S.H. and F.T.) assessed the titles and abstracts of the remaining studies. Ultimately, two authors (S.H. and S.S.) separately evaluated the studies based on their full text. Any disagreement between the two authors was resolved through consensus and subsequently validated by a third author (R.G.).

Data collection process

Data from each study were independently extracted by two researchers (S.H. and S.S.) and entered into a designated "Data Extraction Form." Discrepancies between the two authors were resolved through consensus and subsequently reviewed by a third author (F.T.).

Data items

Data extraction for each included study was conducted according to the following criteria: author, year of publication, sample size, collagen features (defined by the authors of the included studies), collagen feature category (defined by the authors of the present systematic review, including quantity, morphology, microstructural properties, organization, angle related to the tumor boundary, and invented prognostic indices), and association of collagen features with survival, metastasis, grade, stage, therapy response, recurrence, and hormonal receptor (PR, ER, HER2) status.

Quality assessment

Study quality was assessed by two researchers (S.H. and F.T.) using the Newcastle-Ottawa Scale [41] tool. The NOS tool is structured around three domains—selection, comparability, and exposure or outcome—and assigns a total possible score ranging from to 0–9 [42]. Based on this evaluation, the studies were categorized into three levels of quality: good, fair, and poor. In cases of disagreement between the two authors, a third author (Z.M.) was consulted to reach consensus.

Protocol and registration

This systematic review has been registered in the Open Science Framework (OSF), with the URL: OSF | Association of tumoral collagen topology with breast cancer patients' clinical outcomes: A systematic review.

Result

Study selection

A comprehensive search of four major databases (PubMed/MEDLINE, Web of Science, Scopus, and Embase) from inception to September 1, 2024, yielded 21,531 references. After removing duplicates and excluding review articles, 13,081 studies remained for title and abstract screenings. Of these, 12,081 were excluded based on the eligibility criteria. Full-text assessment of the remaining 100 studies led to the exclusion of 56 studies, primarily due to insufficient data or noncompliance with the inclusion criteria. Finally, 44 studies were included in the qualitative analysis. A detailed flowchart illustrating the search process is shown in Fig. 3.

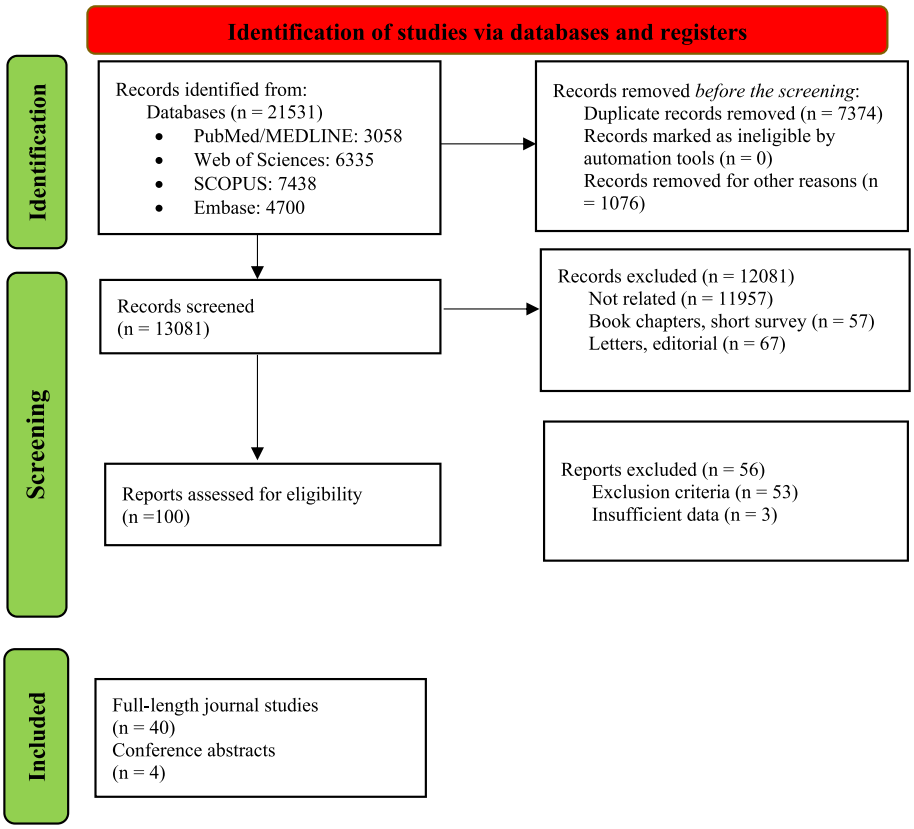

Fig. 3 Flow-chart for the search strategy according to the Preferred Reporting Items for Systematics Reviews and Meta-Analyses (PRISMA) guideline

Study characteristics

All the studies (n = 44) were performed on female human breast tumor samples. Geographically, most clinical studies were conducted in the USA (n = 16) and China (n = 16). The remaining 12 studies were conducted in Singapore (2), Greece (2), Ukraine (2), Norway (1), Brazil (1), United Kingdom (1), Russia (1), Lithuania (1), and Sweden (1). Data were extracted from the included articles to evaluate and integrate the results (Tables S1 and S2 in Supplementary File 1 and Tables in Supplementary File 2). Noticeably, significant variation has been observed among studies on the type of breast cancer samples, methodology for studying collagen and defining various collagen features. To analyze the association between collagen features and breast cancer outcomes, information about methodology, collagen features, and their categories was designed by the authors of the present paper, evaluated the clinical features of patients, and extracted key findings.

Quality assessment

Of the 40 peer-reviewed studies, the majority (n = 37) were assessed as high-quality according to the NOS quality assessment tool, while the rest (n = 3) were considered fair. Fig. 4 shows the quality of the included studies.

Clinical outcomes

Survival prediction

To describe the prognostic significance of collagen features, data were subdivided into three dedicated sections: (1) collagen quantity, (2) collagen architecture, and (3) collagen-based prognostic indexes. Key findings are summarized in Table 2.

Collagen quantity and survival outcomes The relationship between the quantity of collagen fibers and patient outcomes was studied in multiple studies. These studies suggest that this molecule can have diverse effects depending on molecular subtype of breast cancer and the location of collagen in the TME, such as the intratumor or tumor-nest boundary. In luminal breast cancer, elevated intratumoral collagen quantity correlated with worse overall survival (OS), demonstrated by a hazard ratio (HR) of 7.84 in univariate analysis and significance in multivariate analysis. However, this effect was not consistent across all breast cancer

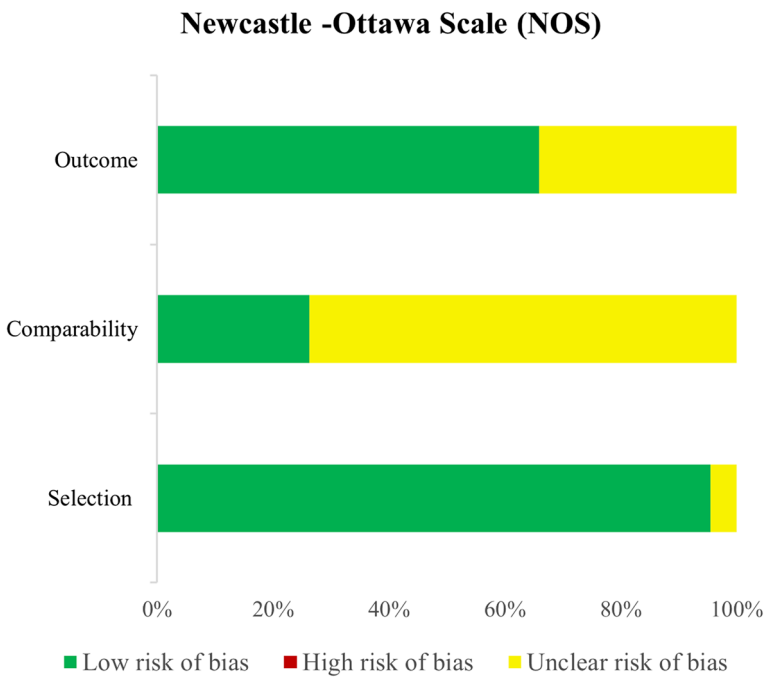

Fig. 4 Summary of quality assessment

subtypes [43]. Two other studies have reported that the total amount of collagen deposition did not correlate with survival [44, 45]. Two studies revealed that a dense layer of collagen fibers surrounding tumor nests was linked to better OS (HR = 0.96) and progression-free survival (PFS, HR = 0.95) [44] and improved disease-free survival (DFS) and OS in triple-negative breast cancer (TNBC) patients [46]. However, two other studies revealed that increased overall collagen density is associated with worse DFS (HR = 2.107) [47] and disease-specific survival (DSS) with no significant correlation [23] in breast cancer patients. Another study revealed that increased intratumoral collagen uniformity (density) did not significantly affect OS [43].

Collagen architecture and survival outcomes Multiple studies have investigated the organization of tumor collagen and its association with patient survival using SHG. One study found that several collagen framework features, including wider collagen fibers (HR=14.25), high variability in collagen fiber width (HR=5.01), a more disorganized arrangement of collagen fibers (HR=4.07), and high variability in collagen fiber arrangement across different regions of the tissue (HR=4.54), were associated with worse overall survival in invasive breast cancer patients, even after adjusting for other factors. On the other hand, median fiber straightness (HR=0.12) was associated with better overall survival [25]. Another study demonstrated that more uniform fibers alignment was associated with worse DFS (HR = 1.96, $p = 0.011$) and that it trended toward reduced DSS (HR = 1.59, $p = 0.10$) [23]. Two studies failed to establish the prognostic value of collagen orientation for OS [27, 43]. The discrepancy between these investigations could be attributed to the angle of the aligned collagen fibers. To clarify, when the studied fibers are oriented parallel to the tumor periphery, more anisotropic structured fibers could be associated with better OS (HR = 0.06, $p = 0.007$) [44]. However, perpendicular angle relative to the tumor boundary with more well-aligned collagen fibers (defined as TACS-3) were significantly associated with worse DFS (HR = 2.59, $p = 0.002$) and DSS (HR = 2.25, $p = 0.008$) compared to less-aligned fibers [23]. Two additional studies reinforced this trend, linking TACS-3 to poorer DFS (HR = 2.39, $p = 0.017$; HR = 3.04, $p = 0.02$) and DSS (HR = 2.17, $p = 0.037$; HR = 3.34, $p = 0.017$) [17, 48]. However, one study found no significant correlation between TACS-3 or combined TACS-2 (aligned collagen constraining tumors) and TACS-3 and OS or progression-free survival (PFS) [44].

Moreover, no significant correlation was discovered between DFS or DSS and fiber curvature [23]. Furthermore, multivariate Cox analysis revealed that better overall survival was marginally related to a lower F/B ratio after adjusting for traditional prognostic factors, although this relationship was not statistically significant ($p = 0.068$) [27].

Table 2 Association of collagen features with survival outcomes

Association	Collagen Feature	Survival Type: HR (*p*-value)	invasive/in situ	Reference
Decreased Survival Rate	Intratumoral Collagen Quantity	OS: HR = 7.84 ($p = 0.031$)	Invasive	[43]
	High Collagenous Stroma	DFS: HR = 2.107 ($p = 0.037$)		[10]
	Collagen Texture Correlation	OS: HR = 4.54 ($p = 0.003$)		[25]
	Mean Fiber Width	OS: HR = 14.25 ($p = 0.010$)		
	Standard Deviation of Fiber Width	OS: HR = 5.01 ($p = 0.011$)		
	Standard Deviation of Magnitude	OS: HR = 4.07 ($p = 0.002$)		
	Fiber Alignment	DFS: HR = 1.96 ($p = 0.011$)		[23]
	TACS-3	DFS: HR = 2.59 ($p = 0.002$) DSS: HR = 2.25 ($p = 0.008$)		
		DFS: HR = 2.39 ($p = 0.017$) DSS: HR = 2.17 ($p = 0.037$)		[17]
		DFS: HR = 3.04 ($p = 0.02$) DSS: HR = 3.34 ($p = 0.02$)		[48]
Increased Survival Rate	Local Collagen Density	OS & DFS	NA	[46]
		PFS: HR = 0.95 ($p = 0.026$) OS: HR = 0.96 ($p = 0.047$)	Invasive	[44]
	Local Collagen Alignment	OS: HR = 0.05 ($p = 0.010$)		
	Median Fiber Straightness	OS: HR = 0.12 ($p < 0.001$)		[25]
No Significant Association With Survival Rate	Intratumoral Collagen Uniformity	OS	Invasive	[43]
	Type I Collagen Protein Expression	OS & BCSS		[45]
	Fiber Density	DSS		[23]
	Total Collagen Deposition	OS & PFS		[44]
	Fiber Curvature	DFS & DSS		[23]
	Intratumoral Collagen Organization	OS		[43]
	TACS-3	OS & PFS		[44]
	TACS-2 and TACS-3			
	Ln F/B	OS	NA	[27]
	Ln R2 value			

Definitions: Collagen Texture Correlation (Higher values indicate more uniform and consistent collagen patterns). Standard Deviation of Magnitude (Higher values reflect greater disorganization in collagen arrangement). Intratumoral Collagen Uniformity (Higher values mean the fibers are more evenly distributed and closely packed). Ln R2 value: (higher values reflect greater anisotropy). Ln F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation)

OS Overall Survival, *DFS* Disease-Free Survival, *BCSS* Breast Cancer-Specific Survival, *DSS* Disease-Specific Survival, *PFS* Progression-Free Survival, *HR* Hazard Ratio, *NA* Not Available, *TACS* Tumor-associated collagen signature, *Ln R2 value* Natural logarithm of the R-squared value, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio

Prognostic value of collagen-based indexes Recent advances in developing prognostic biomarkers in breast cancer leverage TME collagen architecture. Xi et al. [35] developed a model using eight (TACS-1–8), where collagen alignment at the tumor's invasive front predicted outcomes. For example, TACS-1 and TACS-4 (coefficients = 1.575 and 1.101) correlated with better survival, while others, particularly TACS-3 and TACS-6 (coefficients = 1.905 and 1.535) signaled higher risk. When combined with clinical factors, this model achieved strong accuracy for 5-year DFS (AUC: 0.854) and stratified low- and high-risk patients (HR = 5.548) [35]. Complementing this, Xi et al. [26] analyzed TACS corresponding microscopic features (TCMFs), including fiber texture and density, refining predictions. Combining TCMFs with TACS and clinical factors boosted accuracy (AUC: 0.882) and stratified low- and high-risk patients (HR = 6.9) [26]. To capture hidden prognostic clues, Qiu et al. modeled spatial collagen interactions using an intratumor graph neural network (IGNN), which outperformed traditional clinical factors (HR = 8.77) by integrating regional biomarker relationships [36]. At tumor boundaries, Huang et al. identified collagen signatures at tumor boundary (CSTB) that surpassed clinical models in risk stratification (AUC: 0.753) and enhanced nomogram accuracy [49]. Furthermore, four other studies successfully developed unique models centered on collagen features to predict survival [23, 25, 37, 46]. These studies highlight collagen's role as a prognostic marker, offering tools to personalize treatment by identifying undertreated high-risk patients.

Table 3 Association of collagen features with recurrence

Association	Collagen Feature	HR (*p*-value)	Invasive/in situ	Reference
Increased Risk of Recurrence	Intratumoral Collagen Quantity	HR = 4.13 ($p = 0.023$)	Invasive	[43]
	Intratumoral Collagen Uniformity	HR = 3.20 ($p < 0.001$)		
	Ln F/B	HR = 3.39 ($p = 0.019$)		[27]
	Distance to Nearest 2 Fibers	HR = 1.47 ($p = 0.02$)	*In situ*	[24]
	Straightness	HR = 1.47 ($p = 0.03$)		
	Low COL1A1 Protein Expression	Log-rank $p = 0.046$	NA	[10]
Decreased Risk of Recurrence	Total Length Minus End-to-end Length (Curviness)	HR = 0.77 ($p < 0.001$)	*In situ*	[24]
	Width (μm)	HR = 0.57 ($p = 0.01$)		
	Collagen Numbers (within the box with a specific size, 48×48 μm)	HR = 0.58 ($p = 0.003$)		
	Collagen Numbers (within the box with a specific size, 96×96 μm)	HR = 0.60 ($p = 0.004$)		
No Significant Association	Intratumoral Collagen Organization	NS	Invasive	[43]
	Presence of TACS-3 (%)		*In situ*	[50]
	Collagen Alignment			[50]
	Total Length (μm), End-to-end Length (μm), Collagen Numbers (within the box with a specific size, 24×24 μm), Alignment of the Nearest 4 Fibers, Box Alignment 24×24 μm, Distance to Tumor/Stromal Boundary (μm), Angle Relative to Tumor/Stromal Boundary (°)			[24]
	COL1A1 mRNA Expression		NA	[10]

Definitions: Intratumoral Collagen Uniformity (Higher values indicate fibers are more closely packed and evenly distributed). Ln F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation).

NS Not Significant, *NA* Not Available, *TACS* Tumor-associated collagen signature, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio

Recurrence

According to multiple studies, many aspects of collagen fibers are associated with the risk of recurrence (Table 3). One study showed that DCIS cases with wider collagen fiber width and higher collagen numbers (specifically within the box with a specific sizes of 48x48 μm and 96x96 μm) had a better recurrence-free survival. Conversely, DCIS cases with higher fiber straightness and greater distance to the nearest two fibers showed worse recurrence-free survival. Researchers have also introduced five composite factors, three of which are inversely associated with recurrence. These include Factor 2 (high density, long distance to boundary), which is associated with better recurrence-free survival. Factor 3 (serpentine, not straight) is also inversely associated with recurrence. Finally, Factor 5 (long, wide fibers) is also associated with better recurrence-free survival. Factor 1 (high density, low alignment) and Factor 4 (high alignment, parallel to boundary) showed no significant association with recurrence [24]. One study demonstrated that in univariate analysis, elevated intratumoral collagen quantity was associated with a fourfold increased risk of recurrence (HR = 4.13) in luminal breast cancer subtypes. Higher intratumoral collagen uniformity (reflecting fiber density) was linked to a threefold increased relapse risk (HR = 3.20), which remained significant in multivariate models

Table 4 Association of collagen features with invasiveness

Outcome	Collagen Feature	*p*-value	Reference
Negative Association with Invasiveness	Collagen Directionality	$p < 0.0001$	[22]
	Collagen Dispersion	$p = 0.005$	
	F/B ratio in Ductal Carcinoma	$p < 0.05$	[54]
	Collagen Density	$p < 0.0001$	[55]
	Collagen Orientation	$p < 0.001$	[28]
No Association with Invasiveness	Collagen Fiber Quantity	NS	[53]
	Collagen Fiber Organization		
	Collagen Fiber Uniformity		
	Collagen Solidity		[22]
	F/B ratio in Lobular Carcinoma		[54]

Definitions: Intratumoral Collagen Uniformity (Higher values indicate fibers are more closely packed and evenly distributed). F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Collagen Dispersion (Higher values indicate a more random and less parallel arrangement of collagen fibers). Collagen Directionality (Higher values indicate a more perpendicular angle of collagen fibers relative to the boundary). Collagen Solidity (Higher values indicate greater collagen density)

NA Not Available, *NS* Not Significant, *F/B* Forward-to-Backward ratio

intratumoral collagen organization showed no association [43]. Conversely, low expression of COL1A1 protein (indicating reduced collagen synthesis) was linked to inferior 3-year relapse-free survival (RFS; log-rank $p = 0.046$) in young breast cancer patients, though mRNA levels showed no association ($p = 0.781$) [10]. Another

study identified the ln F/B ratio as prognostic marker in ER-positive, distant metastasis-positive patients treated with tamoxifen. Elevated ln F/B correlated with a 3.39-fold higher recurrence risk (HR = 3.39) [27]. Additionally, researchers have observed a positive but insignificant relationship between the alignment of collagen fibers at the tumor boundary of DCIS lesions and recurrence. Moreover, the TACS-3 pattern at the tumor boundary was not correlated with recurrence [50]. However, consistently high and low recurrence rates were associated with high and low TACS (1–8) scores, respectively [35]. Furthermore, various textural features of stromal collagen adjacent to DCIS are significantly associated with DCIS recurrence [51]. Another study introduced a new indicator called the collagen prognostic index. This index is computed by combining three distinct scores: collagen directionality score (rated 1–3), collagen solidity score (rated 1–3), and collagen dispersion score (rated 1–3). These scores are derived from the specific characteristics of collagen fibers, such as their directionality, dispersion, and density. Directionality and dispersion were measured in degrees (°), while solidity was measured in pixels. A high index was an independent predictor of poor prognosis for recurrence [22].

Cancer invasiveness

Five studies analyzed the correlation between breast cancer cell invasiveness and collagen features (Table 4). One study reported a negative relationship between collagen density and breast cancer invasiveness [52], whereas the other reported that the proportion of collagen fibers relative to tumor cells was significantly greater in invasive breast cancer cases than in in situ breast carcinomas [53], and there was no significant relationship between collagen quantity and invasiveness in the two assays [22, 53]. This discrepancy could be due to differences in

Table 5 Association of collagen features of invasive breast tumors with overall stage

Association	Collagen Feature	*p*-value	Comparison	Reference
Positive Association with Overall Stage	Intratumoral Collagen Uniformity	$p = 0.004$	III > I-II	[43]
	Collagen Content	$p = 0.005$	II-IV > 0-I	[47]
No Association with Overall Stage	Total Collagen Deposition Local Collagen Density Local Collagen Alignment TACS-3 TACS-2 + TACS-3	NS	NS	[44]
	Ln F/B	NS	NS	[44]
	Collagen SHG Area Fraction Angular Anisotropy	NS	NS	[57]

Definitions: Intratumoral Collagen Uniformity: (Higher values indicate fibers are more closely packed and evenly distributed). F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Collagen SHG Area Fraction (Higher values indicate a greater collagen content within the tissue). Angular Anisotropy (Higher values indicate collagen fibers are more aligned)

NS Not Significant, *TACS* Tumor-associated collagen signature, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio, *SHG* Second Harmonic Generation

Table 6 Association of collagen features of invasive breast tumors with T stage

Association	Collagen Feature	*p*-value	Comparison	Reference
Positive Association with T Stage	F/B ratio in Lobular Carcinoma	$p < 0.05$	T4 > T1-3	[54]
	Density	$p = 0.003$	T2 > T1	[58]
Negative Association with T Stage	COL1A1 Protein Expression	$p = 0.04$	T1 > T2 > T4 > T3	[10]
	Width	$p < 0.0001$	T1 > T2	[58]
	Length	$p = 0.0006$	T1 > T2	
No Association with T Stage	COL1A1 mRNA Expression	NS	NS	[10]
	F/B Ratio in Ductal Carcinoma	NS	NS	[54]
	Straightness	NS	T1 = T2	[58]
	Angle			

Definitions: F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation)

NS Not Significant, *F/B* the Forward-to-Backward ratio, *COL1A1* Collagen Type I Alpha 1

the methods used to define and calculate the collagen content. Furthermore, one investigation revealed a statistically significant increase in the F/B ratio in ductal carcinoma in situ (DCIS) peritumoral stroma compared with invasive ductal carcinoma (IDC) tumoral stroma; however, this distinction between invasive and in situ carcinomas was not observed in patients with lobular carcinoma. Furthermore, an investigation indicated that the fibers of invasive breast cancers tend to be more perpendicular to the tumor boundary than those of in situ breast cancers [22]. Additionally, one study showed that in patients with lobular carcinoma, collagen alignment is more ordered in in situ breast cancers [28].

Stage

This section delves into the complex relationship between collagen features within invasive breast tumors and the established clinical staging systems. Variations in collagen organization, density, and specific molecular markers are explored in relation to overall stage, tumor size/local extent (T), lymph node involvement (N), and distant metastasis (M).

Overall stage Several studies have explored the link between collagen features and the overall clinical stages of breast cancer (Table 5). In two investigations, higher intratumoral collagen uniformity, indicating more closely packed and evenly distributed fibers, along with increased collagen content, was associated with advanced overall stage [43, 47]. However, another study found no association between F/B ratio and stage [56]. Furthermore, one study revealed no significant correlation between tumor stage and collagen characteristics, including total collagen deposition, local collagen alignment and density, or the presence of TACS-3 and TACS-2+TACS-3 [44]. This was consistent with another study that reported no significant associations between tumor stage and the content or organization of collagen fibers [57].

T component of TNM staging system In the TNM staging system, three papers examined the relationship between collagen features and the T category (Table 6). One study revealed that early T stage breast tissue had a significantly lower F/B ratio than healthy breast tissue. The F/B ratio was significantly elevated in advanced T stage invasive lobular carcinomas (ILCs), but not in IDC patients. These findings suggest that the microstructural characteristics of collagen fibrils during the progression of larger tumors can revert to their initial condition [54]. Another study aimed to assess the expression of collagen type I α1 (COL1A1), which encodes the major component of collagen I, via immunohistochemical (IHC) staining. In the context of T category, there was a gradual decrease in COL1A1 protein expression across T1, T2, and T3 tumors. However, in T4 tumors, COL1A1 expression increased compared to T3 tumors. Notably, the transcript expression of COL1A1 was not significantly correlated with T classification [10]. Furthermore, one study reported that collagen density showed a significant positive association with T stage, with T2 having a higher density than T1, while width and length of collagen fibers showed a significant negative association with T stage, with T1 having a greater width and length than T2. Straightness of collagen fibers showed no significant association with T stage [58].

N and M components of TNM staging system Several studies have explored the connection between collagen characteristics and lymph node status (Table 7) or distance metastasis (Table 8).

Fifteen investigations analyzed the correlation between nodal status and collagen content and density via techniques such as SHG microscopy and staining with Sirius red, Masson's trichrome, or trichrome Mallory (Fig. 5). Collagen content is described in various ways, including the collagen area, density, quantity, uniformity, fiber volume, deposition, intensity, content, and expression. Six of these studies indicated a positive relationship between tumor collagen content and lymph node metastasis [11, 21, 29, 43, 58, 59]. Conversely, nine studies reported no differences in tumor collagen intensity between nodal-positive and nodal-negative groups [10, 20, 29, 43, 44, 47, 57, 60]. Nevertheless, in one case, the authors indicated that statistical significance may be achieved with larger sample sizes [60]. Furthermore, in the context of distance metastasis, two investigations reported a negative correlation between collagen content and distant metastasis [20, 57].

Three studies have evaluated the relationship between lymph node or distance metastasis and the SHG F/B ratio of the tumor stroma, which depends on the fibril diameter and alignment of fibril packing in individual collagen fibers. One study noted that IDC patients with extensive lymph node involvement (N3, $n = 9$) showed a greater F/B ratio compared to those with less involvement (N0–N2, $n = 132$), though the small number of N3 cases might limit this finding. However, another study found no significant difference in the F/B ratio among ILC samples across various lymph node stages (N0–N3) or with distant metastasis (M0 vs. M1). Similarly, no significant F/B ratio differences were observed in IDC samples with varying distant metastatic statuses (M0 vs. M1). Evidence suggests that early in cancer progression, the stromal tissue around tumor cells may exhibit reduced collagen fibril thickness and a more disordered arrangement,

Table 7 Association of collagen features of invasive breast tumors with N stage

Association	Collagen Feature	*p*-value	Comparison	Reference
Positive Association with N Stage	Collagen Alignment	NA	NA	[59]
	Collagen Area	$p < 0.05$		
	COL1A1 Expression	$p < 0.001$	NA	[11]
	Fiber Volume	$p = 0.0036$	NA	[29]
	Co-relation	$p = 0.0007$	NA	
	Collagen Content	$p = 0.015$	NA	[21]
	F/B ratio in Ductal Carcinoma	$p < 0.05$	N3 > N0-N2	[54]
	Collagen Uniformity	$p = 0.034$	NA	[43]
	Density	$p < 0.0001$	N1 > N0	[58]
	Straightness	$p < 0.0001$	N1 > N0	
	Angular Anisotropy	$p = 0.0079$	NA	[57]
Negative Association with N Stage	Eccentricity, Energy	$p < 0.05$	NA	[29]
	Width, Length	$p < 0.0001$	N0 > N1	[58]
No Association with N Stage	F/B ratio in Lobular Carcinoma	NS	NA	[54]
	Collagen Intensity	NS	NA	[60]
	Total Collagen Deposition	NS	NA	[44]
	Local Collagen Density			
	Local Collagen Alignment			
	TACS-3			
	TACS-2 and TACS-3			
	Inter-Fiber Distance	NS	NA	[29]
	Aspect Ratio	NS	NA	
	Angle	NS	N0 = N1	[58]
	Intratumoral Collagen Quantity Intratumoral Collagen Organization	NS	NA	[43]
	Collagen Area Percentage	NS	NA	[20]
	Collagen SHG Area Fraction	NS	N0 = N1	[57]
	COL1A1 Protein Expression	NS	N0=N1-N3	[10]
	COL1A1 mRNA Expression	NS	N0=N1-N3	
	Collagen Content	NS	N0=N1-N3	[47]

Definitions: Intratumoral Collagen Uniformity (Higher values indicate fibers are more closely packed and evenly distributed). F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Collagen SHG area fraction (Higher values indicate a greater collagen content within the tissue). Angular Anisotropy (Higher values indicate collagen fibers are more aligned). Co-relation (Higher values indicate organized spatial patterns). Eccentricity (Higher values indicate parallel alignment). Energy (Higher values indicate homogeneous fiber distribution). Aspect Ratio (Higher values indicate more anisotropic alignment)

Abbreviations: NA Not Available, *NS* Not Significant, *TACS* Tumor-associated collagen signature, *F/B* the Forward-to-Backward ratio, *SHG* Second Harmonic Generation, *COL1A1* Collagen Type I Alpha 1

potentially leading to changes in F/B ratio, with a possibility of reversion to a more original state as the tumor advances [54]. Furthermore, analysis revealed a significant relationship between an increased F/B ratio of the tumor and reduced incidence of metastasis in patients with ER-positive breast cancer [27, 56].

Several studies have focused on the relationship between collagen fiber organization and lymph node metastasis. Three assays revealed that lymph node-positive breast cancers have more organized collagen than lymph node-negative breast cancers [29, 57, 59]. Additionally, the fibers were straighter in lymph node-positive samples [58]. In contrast, one study reported that lymph node-negative breast cancers, compared to lymph node-positive ones, show significantly greater parallel alignment (as indicated by higher eccentricity) and fiber distribution uniformity (as indicated by higher energy). However, there was no significant difference

Table 8 Association of collagen features of invasive breast tumors with M stage

Association	Collagen Feature	*p*-value	Comparison	Reference
Negative Association with M Stage	Ln F/B	$p = 0.004$	NA	[27]
	Ln F/B	$p < 0.005$	NA	[56]
	Collagen Area Percentage	$p = 0.0102$	NA	[20]
	Collagen SHG Area Fraction	$p = 0.0072$	M0 > M1	[57]
No Association with M Stage	F/B Ratio	NS	NA	[54]
	Ln R2 value	NS	NA	[27]

Definitions: F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Ln R2 value: (higher values reflect greater anisotropy). Collagen SHG area fraction (Higher values indicate a greater collagen content within the tissue)

Abbreviations: NA Not Available, *NS* Not Significant, *Ln R2 value* Natural logarithm of the R-squared value, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio

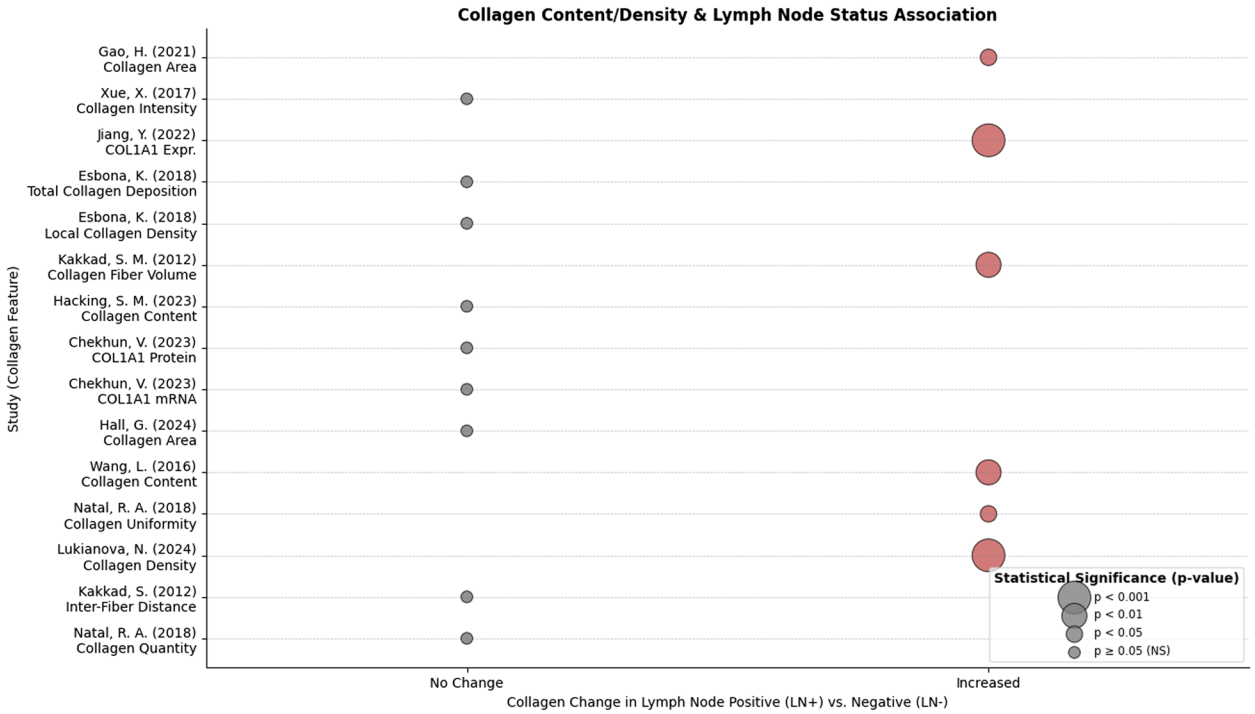

Fig. 5 Scattered plot summarizing the results of studies investigating the association between lymph node status of invasive breast cancer patients with content/density of collagen in the tumor microenvironment. Each dot relates to a reference showing the name of the first author (publishing year), and the collagen content/density-related variable. The size of the dots corresponds to the *p*-value, as shown in the guide box. Collagen Uniformity indicates whether fibers are closely packed and evenly distributed

in anisotropic alignment (as indicated by aspect ratio) between the two groups [29]. Another study found that lymph node-negative samples exhibited significantly greater length and width of collagen fibers than lymph node-positive cases. There was no significant difference in the angle of tumor growth between the two groups [58].

In one study, there was no statistically significant association between MFS length and the anisotropy of the overall orientation of collagen fibers, as measured by fast Fourier transform (FFT) image analysis, within the combined (ER-positive and ER-negative) sample set or in the ER-positive and ER-negative subpopulations, respectively. However, regarding the ER-positive group results, the negative relationship between MFS length and anisotropy ($p = 0.058$) may still be considered questionable and should be examined further [27]. Furthermore, two other studies did not find any association between fiber alignment and lymph node metastasis [43, 44]. Jiang et al. reported that taut,

Table 9 Association of collagen features with grade

Association	Collagen Feature	*p*-value	Comparison	Invasive/*In situ*	Reference
Positive Association with Grade	Total Collagen Deposition	$p < 0.0001$	NA	Invasive	[44]
	Collagen Solidity	$p < 0.0001$	III > II > I	*In situ*	[22]
	Collagen Density	$p < 0.0001$	III > II	NA	[58]
	Collagen Directionality	$p < 0.0001$	III > II > I	*In situ*	[22]
	Anisotropy	$p < 0.0001$	II > I	NA	[15]
	Local Collagen Alignment	$p = 0.014$	NA	Invasive	[44]
	TACS-3	$p = 0.039$	NA		
	TACS-2 and TACS-3	$p = 0.030$	NA		
	Homogeneity, Straightness (annotated by semi-automated methods), Number of Pixels in The Mask (annotated by high detail manual method), Energy, Informational Measure of Correlation 1,	$p < 0.010$	III > I-II	Invasive	[25]
Negative Association with Grade	Collagen Dispersion	$p < 0.0001$	I-II > III	*In situ*	[22]
	COL1A1 Protein Expression	$p = 0.0163$	II > III	NA	[10]
	Collagen Density	$p < 0.05$	Low-grade > high-grade	*In situ*	[62]
	Collagen Density	$p < 0.0001$	Low-grade > high-grade	Invasive	[52]
	Collagen Density	$p < 0.05$	Low-grade > high-grade	*In situ*	[63]
	Collagen Numbers (within the box with a specific sizes, 48x48 µm and 96x96 µm), Distance to nearest 2 fibers, Total length minus end-to-end length, Width	$p < 0.05$	I > II > III	*In situ*	[24]
	F/B ratio	$p < 0.05$	I > II	Invasive	[54]
	Fractal Dimension, Number of Endpoints, Contrast, Inertia, Average, Variance, Entropy, Magnitude, Number of pixels in the mask (annotated by semi-automated method), Straightness (annotated by high detail manual)	$p < 0.05$	I-II > III	Invasive	[25]
	Ratio Parameter Values	$p < 0.0001$	II > III	NA	[15]
	Length	$p < 0.0001$	II > III	NA	[58]
	Width	$p = 0.0042$			
No Association with Grade	Total Length, Straightness, Collagen Numbers (within the box with a specific size, 24x24 µm), Alignment of the Nearest 4 Fibers, Box Alignment 24x24 µm, Distance to Tumor/Stromal Boundary, Angle Relative to Tumor/Stromal Boundary	NS	NS	*In situ*	[24]
	Local Collagen Density	NS	NS	Invasive	[44]
	Collagen Fiber Density	NS	NS	NA	[46]
	The Presence of TACS-3	NS	NS	*In situ*	[50]
	Collagen Fiber Straightness	NA	NS	Invasive	[32]
	Variation in Fiber Direction				
	F/B ratio	NS	I = III II = III	Invasive	[54]
	Ln F/B	NA	NS	NA	[27]
	Collagen Fiber Length	NS	NS	NA	[46]
	COL1A1 mRNA Expression	NS	NS	NA	[10]
	Straightness	NS	NS	NA	[58]
	Angle to the Regions of Tumor Growth				

Definitions: F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Collagen Dispersion (Higher values indicate a more random and less parallel arrangement of collagen fibers). Collagen Directionality (Higher values indicate a more perpendicular angle of collagen fibers relative to the boundary). Collagen Solidity (Higher values indicate greater collagen density). TACS-3: (characterized by bundles of straightened and aligned collagen fibers that are oriented greater than 60 degrees perpendicular to the tumor boundary). TACS-2 and TACS-3: (Refers to the presence of collagen fibers

Table 9 (continued)

oriented at greater than 45 degrees perpendicular to the tumor boundary). Ratio Parameter Values: (Higher value indicates a more intact and symmetrically structured collagen). Homogeneity (Higher values mean smoother texture). High Detail Manual: (Experts create more intricate manual annotations to capture complex collagen patterns). Semi-Automated: (Automated thresholding is combined with manual refinement for more objective annotations). Energy: (Higher values mean more uniform texture). Informational measure of correlation 1: (Higher values mean complex texture with correlated pixel relationships). Fractal dimension: (Higher values mean more complex collagen patterns). Number of endpoints: (Higher values mean a denser, more branched collagen network). Contrast: (Higher values mean more texture variations). Inertia: (Higher values mean more local contrast and complex texture). Average: (Higher values reflect the average intensity of pixel pairs). Variance: (Higher values indicate variability in the sum of pixel pairs). Entropy: (Higher values mean more randomness in collagen texture). Magnitude: (Higher values mean stronger alignment of collagen fibers). Number of pixels in the mask: (Higher values indicate a greater extent of collagen presence in the image)

Abbreviations: NA Not Available, *NS* Not Significant, *TACS* Tumor-associated collagen signature, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio, *COL1A1* Collagen Type I Alpha 1

parallel collagen fibers around tumor nests were the most frequently observed collagen features in a lymph node-negative breast cancer group. The lymph node-positive group had more straightened and aligned collagen fibers, tending to be perpendicular to the tumor boundary [11]. Furthermore, there was a strong association between lymph node metastasis and TACS-1 and −4 (negative), and TACS-5 and −6 (positive). Moreover, they demonstrated that the likelihood of lymph node metastasis increased with increasing TACS (1–8) and TCMF scores [61]. Nevertheless, one investigation revealed no association between collagen alignment and the presence of TACS in different nodal statuses [44].

Grade

Sixteen studies investigated the association between breast cancer grade and collagen features (Table 9). Three investigations have been conducted on the microstructural properties of collagen fibers. One study showed that the F/B ratio of tumor samples of patients with grade I was higher than that of grade II, while there were no significant differences between grades III and I or II [54]; however, the small sample size, especially for grades I ($n = 8$) and III ($n = 15$), may not provide a statistically robust comparison. Another investigation that compared the median F/B ratio showed that there were no differences between different breast cancer grades and F/B ratios [27]. Tsafas et al. showed that, among diverse grades of breast cancer samples, the fiber anisotropy in grade I samples was the lowest, and the presence of fibrils with thin diameters or sparse packing structures in grade III samples was the highest [15].

Among the nine investigations focused on collagen density and tumor grade (Fig. 6), three studies showed no differences in collagen density and different tumor grades [24, 44, 46], while two studies showed a strong positive association between collagen density and tumor

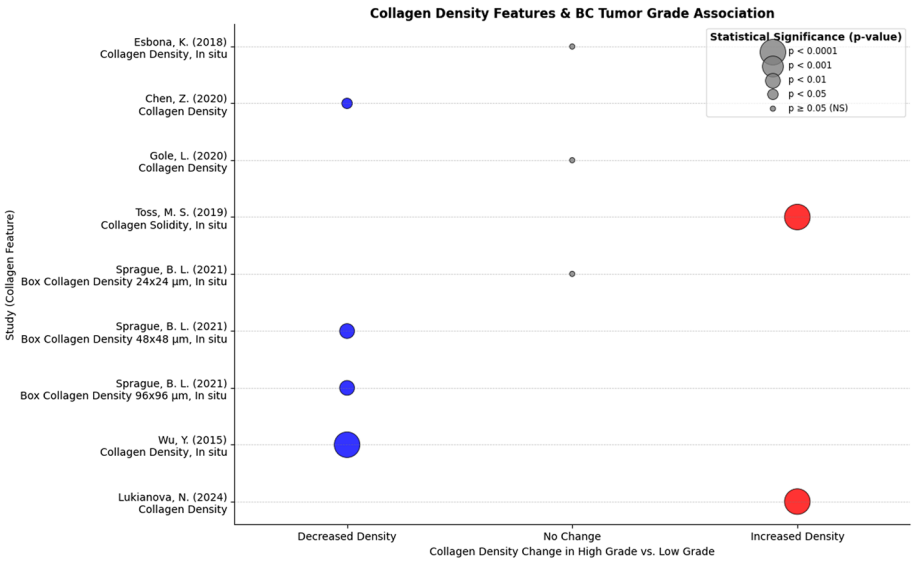

Fig. 6 Scattered plot summarizing the results of studies investigating the association between tumor grade of breast cancer patients with density of collagen in the tumor microenvironment. Each dot relates to a reference showing the name of the first author (publishing year), the collagen density-related variable, and invasive/in situ (indicated where specified by the reference). The size of the dots corresponds to the p-value, as shown in the guide box. Collagen solidity, representing collagen density. 24x24, 48x48, 96x96 indicate counts of fibers within a 24, 48, or 96-micron square box centered on the fiber centroid. BC, breast cancer

grade [22, 58], and four investigations indicated that collagen density in low-grade tumors is significantly higher than that in high-grade tumors [24, 55, 62], underscoring that the relationship between collagen density and tumor grade is not consistently defined across studies and may vary depending on factors such as tumor type or specific methodologies.

Regarding collagen fiber alignment, two studies supported the correlation between tumor grade and increased collagen anisotropy [22, 44], while two other studies did not find any association [24, 32]. Furthermore, when examining the presence of TACS-3, one study found a significant relationship between tumor grade and TACS-3 as well as tumor grade and (TACS-2 + TACS-3) [44]. In contrast, three other studies found no differences [24, 50, 58]. Furthermore, while three studies showed no significant differences between tumor grade and collagen straightness [24, 32, 58], other studies indicated a relationship between higher tumor grade and reduced curviness, width, or length of collagen fibers [24, 58]. Adding to this complexity, there is evidence of a significant positive association between total collagen deposition and tumor grade in an invasive carcinoma study [44], while collagen dispersion was negatively associated with grade (I-II > III) [22]. Similarly, COL1A1 protein expression showed a negative association with grade [47], and distance to the nearest 2 fibers also decreased with increasing grade [24]. Conversely, total length, alignment of the nearest 4 fibers, box alignment (24x24 μm), distance to tumor/stromal boundary, COL1A1 mRNA expression, length and straightness in some studies showed no significant association with tumor grade [24, 32, 46, 47, 58].

In a recent study, researchers developed a novel approach for detecting tumor collagen in bright-field histology images of breast carcinoma, utilizing three distinct annotation methods: Low Detail Manual, where experts subjectively estimate collagen fiber directions with simple lines, resulting in less detailed annotations; High Detail Manual, which involves a more intricate manual annotation by experts to capture complex collagen patterns; and semi-automated, which merges automated thresholding with manual refinement for more objective annotations. The study found that high-grade samples were significantly associated with texture descriptors, such as energy[1], homogeneity[2], and informational measure of correlation 1[3]; however, low-grade samples exhibited higher levels of fractal dimension[4], number of endpoints[5], contrast[6], inertia[7], average[8], variance[9], entropy[10], and magnitude[11]. These descriptors quantify the local variations in the gray-level co-occurrence matrix of the image. Furthermore, the study revealed that straightness and number of pixels in the mask, two key collagen features, varied with tumor grade, depending on the estimation method. High-detail manual annotations showed higher straightness in low-grade samples and more pixels in high-grade samples. In contrast, semi-automated methods indicated higher straightness and pixel count in high-grade and low-grade samples, respectively. This suggests that the annotation method can significantly influence the interpretation of collagen characteristics in breast cancer [25].

Receptor (PR, ER, HER2) status

Seven studies evaluated the relationship between hormone receptors in primary breast cancer cells and collagen structures (Table 10).

According to one study, tumors that were ER^+ and PR^+ had a higher collagen density than those that were negative. However, researchers found no association between HER2 and collagen content [10]. Another study reported that ER^- tumors exhibited significantly higher total collagen deposition and local collagen alignment compared to ER^+ tumors. Additionally, PR^- tumors showed significantly higher local collagen density than PR^+ tumors [44]. According to another experiment, there was no significant relationship between the hormone receptor status of cancer cells and the angle of the fibers relative to the boundary of the tumor. However, the assessment showed

[1] Energy is a texture descriptor derived from the spatial gray-level co-occurrence matrix that measures the uniformity of pixel pair repetitions. High energy values indicate more repetitive patterns and textures in the image.

[2] Homogeneity assesses the closeness of the distribution of elements in the spatial gray-level co-occurrence matrix to the matrix diagonal. Higher homogeneity values suggest less contrast and a more uniform texture across the image.

[3] Informational Measure of Correlation 1 measures the correlation of a pixel to its neighbor over the whole image. It's based on the entropy of the gray-level co-occurrence matrix and the entropy of the image, providing insight into the complexity and the amount of information in the image's texture.

[4] Fractal dimension indicates the complexity of the collagen framework, with higher values suggesting a more intricate and detailed pattern.

[5] A higher number of endpoints typically indicates a more complex and denser collagen network.

[6] A higher contrast value indicates greater disparity in intensity between neighboring pixels, which often corresponds to a more pronounced texture in the image.

[7] Inertia is a method of examining the spatial relationship between pixels. The higher the inertia value, the greater the local contrast and the more complex the texture pattern in the image.

[8] The average intensity value of pixels in the gray-level co-occurrence matrix which reflects the average level of gray in the texture.

[9] Variance measures the spread of the intensity values around the average, which indicates the amount of variation or dispersion from the average.

[10] Entropy is a measure of randomness or complexity in the image texture. High entropy values indicate more complex and less predictable textures.

[11] Magnitude summarizes the collagen framework's directional properties.

Table 10 Association of collagen features with receptor status (ER, PR, HER2)

Collagen Feature	ER Association	PR Association	HER2 Association	Invasive/*In situ*	Reference
Ln F/B	$ER^- > ER^+$ ($p < 0.001$)	$PR^- > PR^+$ ($p = 0.003$)	NS	NA	[27]
TACS-3	$ER^- > ER^+$ ($p = 0.002$)	$PR^- > PR^+$ ($p = 0.02$)	$HER2^+ > HER2^-$ ($p = 0.002$)	*In situ*	[50]
Total Collagen Deposition	$ER^- > ER^+$ ($p = 0.005$)	NS	NS	Invasive	[44]
Local Collagen Density	NS	$PR^- > PR^+$ (p = 0.038)	NS		
Local Collagen Alignment	$ER^- > ER^+$ ($p = 0.031$)	NS	NS		
TACS-3	NS	NS	NS		
TACS-3 and TACS-2	NS	NS	NS		
Ln F/B	NA	NS	NS	Invasive	[56]
Collagen Solidity	$ER^- > ER^+$ ($p = 0.001$)	$PR^- > PR^+$ ($p < 0.0001$)	$HER2^+ > HER2^-$ ($p < 0.0001$)	In situ	[22]
Collagen Dispersion	$ER^+ > ER^-$ ($p = 0.019$)	$PR^+ > PR^-$ ($p = 0.029$)	NS		
Collagen Directionality	NS	NS	NS		
COL1A1 Protein Expression	$ER^+ > ER^-$ ($p = 0.0011$)	$PR^+ > PR^-$ ($p < 0.0001$)	NS	NA	[10]
COL1A1 mRNA expression	$ER^+ > ER^-$ ($p = 0.001$)	$PR^+ > PR^-$ ($p < 0.004$)	NS		
Type I Collagen Protein Expression	$ER^+ > ER^-$ ($p < 0.01$)	NA	NA	Invasive	[45]

Definitions: F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Collagen Dispersion (Higher values indicate a more random and less parallel arrangement of collagen fibers). Collagen Directionality (Higher values indicate a more perpendicular angle of collagen fibers relative to the boundary). Collagen Solidity (Higher values indicate greater collagen density). TACS-3: (characterized by bundles of straightened and aligned collagen fibers that are oriented greater than 60 degrees perpendicular to the tumor boundary). TACS-2 and TACS-3: (Refers to the presence of collagen fibers oriented at greater than 45 degrees perpendicular to the tumor boundary)

Abbreviations: NA Not Available, *NS* Not Significant, *TACS* Tumor-associated collagen signature, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio, *COL1A1* Collagen Type I Alpha 1, *ER* Estrogen receptor, *PR* Progesterone Receptor, *HER2* Human Epidermal Growth Factor Receptor 2

Table 11 Association of collagen features with therapy response

Association	Collagen Feature	*p*-value	Therapy	Invasive/*In situ*	Reference
Negative Association with Therapy Response	Ln F/B	$p = 0.019$	Tamoxifen	NA	[27]
	Tortuosity (1–4), Angle (1–2), and Width (1–2)	$p < 0.05$	Chemotherapy (Paclitaxel) + HER2-Targeted Agents (Trastuzumab/Pertuzumab)	NA	[66]
	Collagen Content	$p = 0.005$	Multi-therapy (Adriamycin/Cytoxan, Paclitaxel, Carboplatin, Bevacizumab)	Invasive	[47]
	Collagen Volume	$p < 0.001$	NA	Invasive	[65]
Positive Association with Therapy Response	F/B (1)	$p = 0.0246$	NACT (a taxane, an anthracycline, cyclophosphamide, etoposide, 5-fluorouracil, methotrexate, and/or platinum therapy) + Mastectomy	NA	[64]
No Association with Therapy Response	F/B (2)	NS			
	Average fiber angle at the tumor boundary				

Definitions: F/B: (Higher value indicates larger fibril diameter, increased order in interfibrillar packing, and decreased fiber segmentation). Tortuosity 1: (90 th percentile tortuosity of fibers in tumor nest border). Tortuosity 2: (standard deviation of fiber tortuosity in tumor nest borders). Tortuosity 3: (mean tortuosity of fibers in tumor nest borders). Tortuosity 4: (median fiber tortuosity in densely inflamed stroma). Angle 1: (standard deviation of fiber relative-angles in bulk tumor borders). Angle 2: (proportion of fibers with high relative-angles in bulk tumor borders). Width 1: (kurtosis of fiber widths in tumor nest borders). Width 2: (skewness of fiber widths in tumor nest borders). F/B 1: (F/B in the tumor host interface of HER2+ patients, selecting pixels was based on an intensity threshold). F/B 2: (F/B in Triple-negative breast cancer patients)

Abbreviations: NA Not Available, *NS* Not Significant, *TACS* Tumor-associated collagen signature, *Ln F/B* Natural logarithm of the Forward-to-Backward ratio, *NACT* Neoadjuvant Chemotherapy

that low collagen density was detected more in ER^-, PR^-, or $HER2^+$ and high dispersion was seen in tumors whose cancer cells were ER^+ or PR^+ [22].

For the determination of microstructural properties of breast cancer with different hormone receptor statuses, one study reported no association [56]. In contrast, another assay showed a higher F/B ratio in ER^- and PR^- -tumors than in ER^+ and PR^+ -tumors [27]. Finally, two articles delved into the relationship between hormone receptor status and the presence of different TACS. In

this regard, while one study showed no significant relationship [44], another study confirmed that TACS-3 was higher in ER^-, PR^-, and $HER2^+$ [50].

Therapy response

Five papers focused on the relationship between collagen features and therapeutic response (Table 11). According to one study, individuals with metastatic ER+ breast cancer, who were treated with tamoxifen, exhibited poorer progression-free survival (PSF) if their pretreated primary tumor had a higher F/B ratio than did those with a lower F/B ratio in the pretreated primary tumor. [27]. In HER2+ biopsy samples of pretreated primary tumors, a higher F/B ratio at the tumor-host interface has been linked to a better pathological response following neoadjuvant chemotherapy (NACT) and tumor resection, as determined by the residual cancer burden (RCB) classification. However, the F/B ratio was not significantly correlated with RCB class in TNBC biopsies [64]. Two studies reported a negative association between collagen content and neoadjuvant therapy response [47, 65]. One study revealed that fibers perpendicular to the tumor boundary tangents, fiber tortuosity, and thickness in the tumor nest borders of specimens from stage II-III HER2-positive breast cancer patients were negatively associated with pathological complete response (pCR) [66]. Another study revealed that the average fiber angle at the tumor boundary was not correlated with RCB class in HER2-positive or TNBC biopsy samples [64].

Discussion

This systematic review comprehensively examined the association between collagen architecture and breast cancer patient outcomes through September 2024. To understand this complex relationship, we explored multiple aspects of breast cancer, including survival, metastasis, treatment response, recurrence, tumor grade, stage, and receptor status. Many studies have employed advanced techniques such as SHG to study fibrillar collagen. Building on this foundation, we highlighted the potential of TACS scores and other computational models incorporating collagen features as prognostic tools. Additionally, by comparing the findings from multiple studies, we investigated the relationship between specific collagen components and distinct breast cancer characteristics, revealing potential clinical implications, conflicting results, and knowledge gaps.

While the majority of studies consistently revealed that specific collagen characteristics (including increased collagen quantity, a more aligned and anisotropic orientation, fibers arranged perpendicular to tumor boundaries, poorer microstructural properties (lower F/B SHG ratio), and straighter fibers) correlated with poorer prognosis in human breast cancer, substantial investigations found no association, and a smaller group of studies reported opposite results. Mechanistically, the increased stiffness of the collagen-rich ECM can impede immune cell infiltration [67] while facilitating cancer cell migration [68] and altering cellular metabolism to favor tumor cell invasion [69]. Conversely, collagen can play a role in antitumor immunity by providing a scaffold for antitumor cytokines [70]. However, induction of T-cell exhaustion through collagen-mediated leukocyte-associated immunoglobulin-like receptor 1 (LAIR-1) signaling complicates this picture [71]. Additionally, collagen may contribute to tumorigenesis by internalizing into cancer cells, providing energy and amino acids for their growth in a typically hypoxic and nutritionally stressed TME [72]. Thus, the overall impact of collagen on cancer progression depends on the delicate balance between these opposing effects within the complex tumor microenvironment. Additionally, aligned collagen fibers create highways that facilitate tumor cell migration and spread, especially when they are arranged perpendicular to the tumor boundary [73]. The collagen receptor, discoidin domain receptor 1 (DDR1), is crucial for establishing an aligned collagen architecture. The DDR1 ectodomain cleavages and remodels the collagen matrix into a parallel structure [74]. These aligned collagen fibers in perivascular regions and around tumor epithelial cells dictate the migratory trajectory of T cells and restrict them from entering tumor islets [41]. Anti-DDR1 antibodies disrupt collagen alignment and facilitate immune cell infiltration [74].

Based on the elegant studies by Burke, K. [27, 54] and Dessa, D. [56, 64], the poorer microstructural properties of collagen, as indicated by a lower F/B SHG ratio, are significantly associated with unfavorable prognosis in human breast cancer. This observation aligns with findings in pancreatic cancer, where increased matrix metalloproteinase (MMPs)-mediated cleaved COL-I (cCOL-I) has been linked to tumor progression [75]. This indicates that a low F/B SHG ratio (reflecting a predominance of more segmented and thinner collagen fibrils with more random inter-fibril packing structures) [12]) might indicate a higher proportion of cCOL-I within the tumor microenvironment. cCOL-I engages DDR1, initiating a signaling cascade that stimulates collagen macropinocytosis and mitochondrial biogenesis, enhances tumor bioenergetics, and ultimately promotes tumor growth and metastasis. Therefore, a lower F/B SHG ratio, as hypothesized by the higher abundance of cCOL-I, could contribute to a more aggressive tumor phenotype. Furthermore, increased metabolic activity driven by cCOL-I-mediated

signaling might create a more permissive environment for tumor cell proliferation and invasion [75, 76].

However, the results of studies examining the relationship between tumor-stromal collagen characteristics and patient outcomes in breast cancer are remarkably contradictory, which underscores the complex and multifaceted role of collagen in breast cancer, highlighting the importance of context in determining the effect of collagen and the need for further investigation. To fully harness the potential of this research area, it is crucial to establish standardized methods for analyzing collagen features. By implementing standardized methodologies, researchers can better understand the influence of collagen on cancer progression. This will not only facilitate the integration of different studies but also pave the way for the development of cost-effective and valuable collagen-based prognostic tools for clinical use.

To fully understand the complex relationship between TME collagen and patient outcomes, future studies should address several critical factors.

First, ensuring an adequate sample size across all study groups is crucial. Smaller studies, as seen in some study groups of the reviewed research [27, 54] are more susceptible to chance variations and may not represent a broader patient population. Increasing sample sizes strengthens the validity of the findings and allows for more generalizable conclusions [77].

Second, standardizing methodologies for collagen analysis is highly beneficial. This encompasses various aspects, such as tissue processing and staining approaches, detection techniques, and software and calculation methods used for image analysis [78]. For instance, various steps in the processing and mounting of sample tissues may affect the F/B ratio, emphasizing the need for standardized protocols [64]. The current lack of standardization creates challenges when researchers compare the findings of different studies. By establishing standardized techniques, future research can seamlessly integrate data and pave the way for more conclusive results.

Third, addressing the inconsistencies in the definition and measurement of collagen features is critical. Our review highlighted discrepancies, such as the term "uniformity" encompassing fiber density or distribution in various studies [43, 53]. Similar inconsistencies exist for other features, such as alignment [23, 44, 50, 59] and density [22–25, 43, 44, 46, 52, 53, 55, 59, 62]. This lack of standardization hinders our ability to compare findings and ultimately to understand how collagen influences patient outcomes. Establishing consistent definitions and measurement techniques is the key to unlocking the full potential of this research area.

Fourth, determining the spatial distribution of collagen fibers relative to the tumor is critical. In situ breast cancer primarily features peritumoral collagen, whereas invasive breast cancer exhibits both peritumoral and intratumoral collagen, with collagen often interspersed among cancer cells at the invasive front [34, 35]. Notably, distinct collagen characteristics within these spatial contexts may differentially affect disease outcomes. For instance, while total collagen deposition in invasive breast cancer did not correlate with survival, local collagen density, measured as collagen content within a 50-pixel radius from the tumor-normal boundary, was significantly associated with a better prognosis [44]. This discrepancy, potentially attributable to variations in the collagen region of interest and methodological approaches, suggests that collagen quantity alone may not fully capture its influence on tumor behavior.

Fifth, the clinical heterogeneity of breast cancer, including receptor expression, tumor grade, invasiveness, origin, treatment history, and age, is crucial for interpreting collagen-related findings. The prognostic value of collagen is context-dependent, as evidenced by disparate results across studies. For instance, while an elevated F/B ratio correlates with favorable outcomes in ER+ breast cancer patients, its prognostic significance in other subpopulations remains equivocal [27]. Given that COL1A1 expression is often higher in ER+ tumors [10, 45], potentially influenced by estrogen, and F/B (reflecting fibril diameter and packing) was observed to be lower in ER+ patients [27], it could be hypothesized that while estrogen may promote collagen production, it might also contribute to disrupted fibrillogenesis within the malignant context. Meanwile, in ER+ group, patients exhibiting a higher F/B ratio might represent a subgroup where fibril diameter and packing are relatively preserved, potentially counteracting some of the adverse effects associated with estrogen-driven ECM alterations and leading to improved outcomes. One study reported a positive association between F/B ratio in ductal carcinoma and N stage, with N3 tumors showing a higher F/B than N0-N2. This is particularly interesting as N3 represents more extensive lymph node metastases. The authors speculate this "normalization" of F/B in highly metastatic tumors might facilitate further spread. Intensity threshold-based F/B at the tumor-host interface is correlated with the RCB class in HER2+ biopsy samples, but this correlation was not detected in TNBC samples [64]. Moreover, a notable study reported a higher F/B ratio in DCIS peritumoral stroma collagen than in collagen fibers within the IDC tumor, but this disparity was absent in lobular carcinoma [54]. Additionally, variations in ECM remodeling from early-stage to late-stage disease and differences in mutation profiles further modulate collagen

characteristics, contributing to the complex and sometimes conflicting prognostic associations observed across studies [34]. For example, *TP53* mutations are often associated with increased MMP activity, potentially leading to a more fragmented collagen network [79], while *PIK3CA* mutations might enhance growth factor signaling that promotes collagen deposition and alignment [80]. This emphasis on the complex nature of the relationship between tumor cells and collagen characteristics and underscores the need for more comprehensive and well-designed research. Given the importance of these studies, the authors have provided all available patient data of included papers in Supplementary File 2. These detailed patient data can help researchers develop more accurate prognostic models and targeted therapies based on collagen characteristics.

Conclusion

In conclusion, this review highlights the complex and often contradictory roles of collagen in breast cancer. Although promising, the field requires standardized methodologies and larger studies to establish collagen as a reliable prognostic or predictive biomarker. Given its potential impact on tumor biology, exploring the role of collagen in other cancers is warranted.

Abbreviations

BCSS	Breast cancer-specific survival
CSS	Cancer-specific survival
cCOL-I	cleaved COL-I
CSTB	Collagen signatures at tumor boundary
CIs	Confidence intervals
COL1A1	Collagen type I α1
CVF	Collagen volume fraction
DDR1	Discoidin domain receptor 1
DFS	Disease-free survival
DSS	Disease-specific survival
DCIS	Ductal carcinoma in situ
ECM	Extracellular matrix
F/B	Forward-emitted to backward-emitted
HRs	Hazard ratios
H&E	Hematoxylin and eosin
IHC	Immunohistochemistry
IGNN	Intratumor graph neural network
IDC	Invasive ductal carcinoma
LAIR-1	Leukocyte-associated immunoglobulin-like receptor 1
MMP	Matrix metalloproteinase
MFS	Metastasis-free survival
NACT	Neoadjuvant chemotherapy
NOS	Newcastle–Ottawa scale
OS	Overall survival
pCR	Pathological complete response
PSF	Progression-free survival
RFS	Relapse-free survival
RCB	Residual cancer burden
SHG	Second harmonic generation
TCMF	TACS corresponding microscopic features
TME	Tumor microenvironment
TACS	Tumor-associated collagen signature
WOS	Web of Science

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-025-02017-6.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements

Not applicable.

Authors' contributions

R.G. and S.H. conceived and designed the study. F.T. provided the search syntax. F.T. conducted the database search. S.H., F.T., S.S., and S.D. screened and selected data. F.T. assessed data quality. S.H. extracted data and prepared the initial draft of the manuscript, while F.T. drafted the Method section. S.H. prepared and illustrated Figures and Tables. F.K. and B.K. provided expert opinions. Z.M. and R.G. reviewed and edited the final manuscript. All authors contributed to data interpretation and critical revision of the manuscript for important intellectual content.

Funding

This study was funded by a grant from the Iran University of Medical Sciences (Grant No: 1403-2-99-30621).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This systematic review was approved by the Ethics Committee of the Iran University of Medical Sciences (IR.IUMS.REC.1403.545). As this review relied solely on publicly accessible published data, formal consent to participate was not required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Oncopathology Research Center, Iran University of Medical Sciences, Tehran, Iran. [2]Department of Molecular Medicine, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran 14496-14530, Iran. [3]Department of General Surgery, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [4]Department of Physics, Sharif University of Technology, Tehran, Iran.

Received: 8 December 2024 Accepted: 4 April 2025
Published online: 05 May 2025

References

1. Bray F, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Anderson NM, Simon MC. The tumor microenvironment. Curr Biol. 2020;30(16):R921–5.
3. Li JJ, Tsang JY, Tse GM. Tumor microenvironment in breast cancer—updates on therapeutic implications and pathologic assessment. Cancers. 2021;13(16):4233.
4. Lepucki A, et al. The role of extracellular matrix proteins in breast cancer. J Clin Med. 2022;11(5):1250.

5. Martinez J, Smith PC. The dynamic interaction between extracellular matrix remodeling and breast tumor progression. Cells. 2021;10(5):1046.
6. Lo Buglio G, et al. The multifaced role of collagen in cancer development and progression. Inte J Mol Sci. 2024;25(24):13523.
7. Zhang Q, et al. Collagen code in tumor microenvironment: functions, molecular mechanisms, and therapeutic implications. Biomed Pharmacother. 2023;166:115390.
8. Cambi A, Ventre M. Collagen-based biomimetic systems to study the biophysical tumour microenvironment. Cancers (Basel). 2022;14(23):5939.
9. Baldari S, et al. Strategies for efficient targeting of tumor collagen for cancer therapy. Cancers. 2022;14(19):4706.
10. Chekhun V, et al. Features of COL1A1 expression in breast cancer tissue of young patients. Exp Oncol. 2023;45(3):351–63.
11. Jiang Y, et al. Collagen fiber features and COL1A1: are they associated with elastic parameters in breast lesions, and can COL1A1 predict axillary lymph node metastasis? BMC Cancer. 2022;22(1):1004.
12. Williams RM, Zipfel WR, Webb WW. Interpreting second-harmonic generation images of collagen I fibrils. Biophys J. 2005;88(2):1377–86.
13. Golaraei A, et al. Polarimetric second-harmonic generation microscopy of the hierarchical structure of collagen in stage I-III non-small cell lung carcinoma. Biomed Opt Express. 2020;11(4):1851–63.
14. Golaraei A, et al. Changes of collagen ultrastructure in breast cancer tissue determined by second-harmonic generation double Stokes-Mueller polarimetric microscopy. Biomed Opt Express. 2016;7(10):4054–68.
15. Tsafas V, et al. Polarization-dependent second-harmonic generation for collagen-based differentiation of breast cancer samples. J Biophotonics. 2020;13(10):e202000180.
16. Oldenbourg R. Polarized light microscopy: principles and practice. Cold Spring Harb Protoc. 2013;2013(11):pdb.top078600.
17. Majeed H, et al. Quantitative histopathology of stained tissues using Color Spatial Light Interference Microscopy (cSLIM). Sci Rep. 2019;9:14.
18. Ding H, et al. Measuring the scattering parameters of tissues from quantitative phase imaging of thin slices. Opt Lett. 2011;36(12):2281–3.
19. Dey P. Basic and advanced laboratory techniques in histopathology and cytology. Singapore: Springer; 2018.
20. Sun C, et al. Quantitative measurement of breast carcinoma fibrosis for the prediction in the risk of bone metastasis. Am J Transl Res. 2018;10(6):1852–9.
21. Wang L, et al. Is macroscopic tumor stiffness on strain elastography related to the axillary nodal status in T1 stage ductal invasive breast cancer? Int J Clin Exp Med. 2016;9(2):3371–9.
22. Toss MS, et al. Geometric characteristics of collagen have independent prognostic significance in breast ductal carcinoma in situ: an image analysis study. Mod Pathol. 2019;32(10):1473–85.
23. Bredfeldt JS, et al. Automated quantification of aligned collagen for human breast carcinoma prognosis. J Pathol Inform. 2014;5(1):28.
24. Sprague BL, et al. Collagen organization in relation to ductal carcinoma in situ pathology and outcomes. Cancer Epidemiol Biomarkers Prev. 2021;30(1):80–8.
25. Morkunas M, et al. Tumor collagen framework from bright-field histology images predicts overall survival of breast carcinoma patients. Sci Rep. 2021;11(1):15474.
26. Xi G, et al. Computer-assisted quantification of tumor-associated collagen signatures to improve the prognosis prediction of breast cancer. BMC Med. 2021a;19(1):273.
27. Burke K, et al. Using second harmonic generation to predict patient outcome in solid tumors. BMC Cancer. 2015;15:929.
28. Shen TF, et al. Monitoring the progression of lobular breast carcinoma using multiphoton microscopy. Laser Physics Letters. 2019;16(10):9.
29. Kakkad SM, et al. Collagen i fiber density increases in lymph node positive breast cancers: Pilot study. J Biomed Opt. 2012;17(11):116017.
30. Nguyen TL, et al. Quantitative phase imaging: recent advances and expanding potential in biomedicine. ACS Nano. 2022;16(8):11516–44.
31. Chen Y, Yu Q, Xu CB. A convenient method for quantifying collagen fibers in atherosclerotic lesions by ImageJ software. 2017.Chen, Ying, Qiuhao Yu, and Cang-Bao Xu. A convenient method for quantifying collagen fibers in atherosclerotic lesions by ImageJ software. Int J Clin Exp Med. 2017;1(10):14904–10.
32. Brabrand A, et al. Alterations in collagen fibre patterns in breast cancer. A premise for tumour invasiveness? APMIS. 2015;123(1):1–8.
33. Perry SW, et al. Stromal matrix metalloprotease-13 knockout alters Collagen I structure at the tumor-host interface and increases lung metastasis of C57BL/6 syngeneic E0771 mammary tumor cells. BMC Cancer. 2013;13(1):411.
34. Provenzano PP, et al. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 2006;4(1):38.
35. Xi GQ, et al. Large-scale tumor-associated collagen signatures identify high-risk breast cancer patients. Theranostics. 2021b;11(7):3229–43.
36. Qiu L, et al. Intratumor graph neural network recovers hidden prognostic value of multi-biomarker spatial heterogeneity. Nat Commun. 2022;13(1):4250.
37. Ren Y, et al. Improved quantitative fibrosis indices reveal diverse survivals of triple negative breast cancer patients. Cancer Res. 2023;83(5):P6-04-11.
38. Provenzano PP, et al. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 2006;4:1–15.
39. Munn Z, et al. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18(1):5.
40. Page MJ, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:n160.
41. Salmon H, et al. Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. J Clin Invest. 2012;122(3):899–910.
42. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25(9):603–5.
43. Natal RA, et al. Collagen analysis by second-harmonic generation microscopy predicts outcome of luminal breast cancer. Tumor Biol. 2018;40(4):1010428318770953.
44. Esbona K, et al. The presence of cyclooxygenase 2, tumor-associated macrophages, and collagen alignment as prognostic markers for invasive breast carcinoma patients. Am J Pathol. 2018;188(3):559–73.
45. Jansson M, et al. Stromal type I collagen in breast cancer: correlation to prognostic biomarkers and prediction of chemotherapy response. Clin Breast Cancer. 2024;24(5):e360-e369.e4.
46. Gole L, et al. Quantitative stain-free imaging and digital profiling of collagen structure reveal diverse survival of triple negative breast cancer patients. Breast Cancer Res. 2020;22(1):13.
47. Hacking SM, et al. Whole slide image features predict pathologic complete response and poor clinical outcomes in triple-negative breast cancer. Pathol Res Pract. 2023;246:154476.
48. Conklin MW, et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am J Pathol. 2011;178(3):1221–32.
49. Huang, Xingxin, Fangmeng Fu, Wenhui Guo, Deyong Kang, Xiahui Han, Liqin Zheng, Zhenlin Zhan, Chen Wang, Qinghua Zhang, Shuhua Wang, and Shusen Xu. Prognostic significance of collagen signatures at breast tumor boundary obtained by combining multiphoton imaging and imaging analysis. Cell Oncol. 2024;47(1):69–80.
50. Conklin MW, et al. Collagen alignment as a predictor of recurrence after ductal carcinoma in situ. Cancer Epidemiol Biomark Prev. 2018;27(2):138–45.
51. Upadhaya T, et al. Prediction of breast ductal carcinoma in situ recurrence using histomics analysis of stromal collagen from second-harmonic generation and hematoxylin and eosin stain-based images. Cancer Res. 2022;82(4):P1-02-16.
52. Wu Y, et al. Monitoring the progression from intraductal carcinoma to invasive ductal carcinoma based on multiphoton microscopy. J Biomed Opt. 2015;20(9):096007.
53. Gubarkova EV, et al. Multiphoton tomography in differentiation of morphological and molecular subtypes of breast cancer: a quantitative analysis. J Biophotonics. 2021;14(5):e202000471.
54. Burke K, Tang P, Brown E. Second harmonic generation reveals matrix alterations during breast tumor progression. J Biomed Opt. 2013;18(3):31106.
55. Wu Y, et al. Monitoring morphological alterations during invasive ductal breast carcinoma progression using multiphoton microscopy. Lasers Med Sci. 2015;30(3):1109–15.
56. Desa DE, et al. Intratumoral heterogeneity of second-harmonic generation scattering from tumor collagen and its effects on metastatic risk prediction. BMC Cancer. 2020;20(1):1217.

57. Hall G, et al. SHG fiberscopy assessment of collagen morphology and its potential for breast cancer optical histology. IEEE Trans Biomed Eng. 2024;71(8):2414–20.
58. Lukianova N, et al. Development of an algorithm for biomedical image analysis of the spatial organization of collagen in breast cancer tissue of patients with different clinical status. FEBS Open Bio. 2024;14(4):675–86.
59. Gao H, et al. 3D extracellular matrix regulates the activity of T cells and cancer associated fibroblasts in breast cancer. Front Oncol. 2021;11:764204.
60. Xue XW, et al. Kindlin-2 could influence breast nodule elasticity and improve lymph node metastasis in invasive breast cancer. Sci Rep. 2017;7:10.
61. Fang Y, et al. Collagen signature as a novel biomarker to predict axillary lymph node metastasis in breast cancer using multiphoton microscopy. J Biophotonics. 2022;15(6):e202100365.
62. Chen Z, et al. Label-free identification of early stages of breast ductal carcinoma via multiphoton microscopy. Scanning. 2020;2020:8.
63. Chen X, et al. Monitoring changes of collagen fibers surrounding breast ductal carcinoma in situ using multiphoton microscopy. In: Proceedings of SPIE - The International Society for Optical Engineering. 2023.
64. Desa DE, et al. Second-harmonic generation directionality is associated with neoadjuvant chemotherapy response in breast cancer core needle biopsies. J Biomed Opt. 2019;24(8):1–9.
65. Huang JX, et al. Comparing shear wave elastography of breast tumors and axillary nodes in the axillary assessment after neoadjuvant chemotherapy in patients with node-positive breast cancer. Radiol Med. 2024;129(8):1143–55.
66. Nguyen TH, et al. Quantitative analysis of fiber-level collagen features in H&E whole-slide images predicts neoadjuvant therapy response in patients with HER2+ breast cancer. Cancer Res. 2023;83(5):P5-02-09.
67. Kuczek DE, et al. Collagen density regulates the activity of tumor-infiltrating T cells. J Immunother Cancer. 2019;7(1):68.
68. Wu Y, et al. Matrix-driven changes in metabolism support cytoskeletal activity to promote cell migration. Biophys J. 2021;120(9):1705–17.
69. Morris BA, et al. Collagen matrix density drives the metabolic shift in breast cancer cells. EBioMedicine. 2016;13:146–56.
70. Momin N, et al. Anchoring of intratumorally administered cytokines to collagen safely potentiates systemic cancer immunotherapy. Sci Transl Med. 2019;11(498):eaaw2614.
71. Peng DH, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8(+) T cell exhaustion. Nat Commun. 2020;11(1):4520.
72. Olivares O, et al. Collagen-derived proline promotes pancreatic ductal adenocarcinoma cell survival under nutrient limited conditions. Nat Commun. 2017;8:16031.
73. Provenzano PP, et al. Collagen density promotes mammary tumor initiation and progression. BMC Med. 2008;6(1):11.
74. Sun X, et al. Tumour DDR1 promotes collagen fibre alignment to instigate immune exclusion. Nature. 2021;599(7886):673–8.
75. Su H, et al. Collagenolysis-dependent DDR1 signalling dictates pancreatic cancer outcome. Nature. 2022;610(7931):366–72.
76. Su H, Karin M. Collagen architecture and signaling orchestrate cancer development. Trends Cancer. 2023;9(9):764–73.
77. Banerjee A, Chaudhury S. Statistics without tears: populations and samples. Ind Psychiatry J. 2010;19(1):60–5.
78. Liu, Yuming, Adib Keikhosravi, Guneet S. Mehta, Cole R. Drifka, and Kevin W. Eliceiri. Methods for quantifying fibrillar collagen alignment. In Fibrosis: Methods and Protocols, edited by Paolo Bonaldo and Mauricio Rojas. New York: Springer; 2017. p. 429–51.
79. Zhu G, et al. Mutant p53 in cancer progression and targeted therapies. Front Oncol. 2020;10:595187.
80. An Y, et al. Cdh1 and Pik3ca mutations cooperate to induce immune-related invasive lobular carcinoma of the breast. Cell Rep. 2018;25(3):702-714.e6.

Publisher's Note