View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repository@Nottingham

Geometric characteristics of collagen have independent prognostic significance in breast ductal carcinoma *in situ*: An image analysis study

Michael S Toss[1,2], Islam M Miligy[1,3], Kylie L Gorringe[4,5], Abdulbaqi AlKawaz[1,6], Karuna Mittal[7], Ritu Aneja[7], Ian O Ellis[1], Andrew R Green[1], Ioannis Roxanis[8] and Emad A Rakha[1,3]

[1]Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham, Nottingham City Hospital, UK. [2]Histopathology department, South Egypt Cancer Institute, Assiut University, Egypt, [3]Histopathology department, Faculty of Medicine, Menoufia University, Egypt, [4]Cancer Genomics Program, Peter MacCallum Cancer Centre, Melbourne, and [5]The Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Australia, [6]College of dentistry, Al Mustansiriya University, Iraq, [7]Georgia State University, Atlanta, GA, USA, and [8]Institute of Cancer Research, London and Royal Free London NHS Foundation Trust, London, UK

Correspondence:

Professor Emad Rakha

Department of Histopathology, Nottingham University Hospital NHS Trust,

City Hospital Campus, Hucknall Road, Nottingham, NG5 1PB, UK

Tel: (44) 0115-9691169, Fax: (44) 0115-9627768

Email: emad.rakha@nottingham.ac.uk, Emad.rakha@nuh.nhs.uk

Key words: DCIS, collagen, poor prognosis, collagen prognostic index recurrence

Running title: Collagen characteristics in DCIS

Abstract

Collagen plays a key role in normal and malignant tissue homeostasis. While the prognostic significance of collagen fibre remodeling in invasive breast cancer has been studied, its role in ductal carcinoma *in situ* (DCIS) remains poorly defined. Using image analysis, we aimed to evaluate the prognostic significance of the geometric characteristics of collagen surrounding DCIS. A large well-characterized cohort of DCIS comprising pure DCIS (n=610) and DCIS co-existing with invasive carcinoma (n=180) were histochemically stained for collagen using picrosirius red. ImageJ software was used to assess collagen density, degree of collagen fibre dispersion and directionality in relation to DCIS ducts' boundary. We developed a collagen prognostic index and evaluated its prognostic significance. A poor index was observed in 24% of the pure DCIS and was associated with determinants of high-risk DCIS including higher grade, comedo necrosis, hormonal receptor negativity, HER2 positivity and high proliferation index. High index was associated with overexpression of the collagen remodeling protein prolyl-4-hydroxlase alpha 2 and the hypoxia inducible factor 1α. DCIS co-existing with invasive carcinoma had a higher collagen prognostic index than pure DCIS (p<0.0001). High index was an independent poor prognostic factor for all DCIS recurrences (HR=2.3, p=0.005) and just invasive recurrences (HR=3.4, p=0.003). Interaction with radiotherapy showed that the higher index was associated with poor outcome even with adjuvant radiotherapy (p=0.0001). In conclusion, collagen re-organization around DCIS is associated with poor outcome and provides a potential predictor for disease progression and resistance to radiotherapy. Mechanistic studies are warranted to decipher the underlying mechanisms.

Introduction

The key difference between invasive breast cancer and ductal carcinoma *in situ* (DCIS) is the destruction of the surrounding basement membrane and invasion of the malignant cells into the surrounding stroma accompanied by desmoplastic reaction. However, studies have shown that changes in the surrounding tumor microenvironment including infiltrating immune cells, stromal cells and extracellular matrix composition occurs early in carcinogenesis much before malignant cells invade the surrounding stroma (1-5). More so, these changes may provide a surrogate to predict tumor behavior and potential outcomes (5-7). One of the major limitations of currently available clinicopathological indices and molecular signatures to predict DCIS risk is the underestimated role of microenvironmental factors. Therefore, despite advances in management modalities, a considerable percentage of DCIS are either over- or under-treated. In addition, most molecular studies that compared the genetic profiles of DCIS and invasive breast cancer report a high degree of similarity (8, 9), creating a compelling need to decipher the role of the tumor microenvironment in DCIS behavior.

Collagen is the main constituent of extracellular matrix and plays a crucial role in tissues homeostasis (10). There are many types of collagen with varied localizations and functions in normal and pathological conditions (2, 10). The role of collagen in cancer development is not solely related to its chemical composition or signaling interaction but is mainly linked to its physical properties and geometric characteristics in relation to the tumor cells (4, 11-13). Recent advances in imaging techniques and specific histochemical staining of collagen have spurred an increased interest to discover its physical roles and biological functions (12, 14-16). In invasive cancers including breast cancer, studies have shown that densely aligned collagen is associated with poor outcome and various tumor-associated collagen signatures have been described to assess the collagen organization surrounding invasive breast cancer (11). There is a dearth of studies deciphering the role of collagen in DCIS; hence more investigations are needed to evaluate its prognostic significance. While an earlier study has shown association of aligned collagen with other determinants of DCIS aggressive behavior, no statistically significant association with disease recurrence or progression was reported (6). In the current study, we hypothesized that collagen re-

organization surrounding DCIS is linked to aggressive behavior. We have developed an imaged-based collagen prognostic index that relies on three factors: the directionality of collagen fibre relative to DCIS boundary, collagen density and dispersion using a large cohort of DCIS.

MATERIAL AND METHODS

Study Cohort

A large well-characterized and annotated cohort of DCIS including pure DCIS (n=776) and DCIS mixed with invasive breast cancer (DCIS-mixed) (n=239) diagnosed between 1990 to 2012 at Nottingham City Hospital, Nottingham, United Kingdom was used (17). To avoid selection bias, the DCIS-mixed cohort was selected with clinicopathological features comparable to the pure cohort regarding age at diagnosis, DCIS nuclear grade, and the presence of comedo type necrosis. All demographic, clinical, histological, management information was collected. Breast conserving surgery was the primary surgical option for 70% of cases, however due to close or positive surgical margins a considerable percentage of patients underwent completion mastectomy; half of the cases were treated by mastectomy as the final operation. Data about the molecular classes, tumor infiltrating lymphocytes density (7), as well as the key regulator of collagen biosynthesis and remodeling prolyl-4-hydroxlases alpha subunit 2 (10) and the hypoxia related marker hypoxia inducible factor 1α were available. In brief, the molecular classes were defined based on the immunohistochemistry using estrogen (ER) and progesterone (PR) receptor, the human epidermal growth factor receptor 2 (HER2) and Ki67 proliferation index. ER and PR were considered positive if >1% of tumor cells showed nuclear staining (18) while HER2 positivity was defined when more than 10% of tumor cells showed strong complete membranous staining (+3 score). Chromogenic *in situ* hybridisation was used to determine the gene amplification status in borderline cases (+2 score) (19). The proliferation index was defined as low when <14% of cells showed nuclear staining for Ki67 (20). Hypoxia inducible factor 1α was evaluated using immunohistochemistry and was considered positive when >1% of tumor cells showed nuclear staining as previously described (21). Dense tumor infiltrating lymphocytes DCIS was defined in DCIS as an average number of ≥20 lymphocytes/duct (7). Follow up data was collected and local recurrence

free interval was defined as the time (in months) between 6 months after the primary DCIS excision and the development of ipsilateral recurrence (either as DCIS or invasive breast cancer). Cases undergoing completion re-excision due to margins' involvement or presence of residual tumor within the first six months of the primary operation were not considered as disease recurrence. Contralateral disease following DCIS diagnosis was not accounted as a recurrence and these patients were censored at the time of the development of the contralateral cancer. Within a median follow up period of 103 months (range 6-240), 83 cases (11%) developed a recurrence in the pure DCIS cohort with 30 DCIS (36%) and 53 invasive recurrence (64%) (10, 22). Most recurrences (n=66) developed in patients treated with breast conserving surgery alone.

Histochemical collagen staining with picrosirius red

Tissue microarrays were prepared from both cohorts as previously described (17). Briefly, 1 mm punch sets were used to construct the tissue microarray using an automated GRAND MASTER 2.4-UG-EN tissue microarray machine. For better representation of cases with heterogeneous DCIS histological patterns and/or nuclear grade, multiple cores were sampled from all representative areas.

Picrosirius red solution was prepared by mixing 0.5 g of Sirius Red [Direct red 80, Cat#365548, Sigma-Aldrich, Darmstadt, Germany] with 500 ml of saturated aqueous solution of picric acid [Cat# P6744-1GA, Sigma-Aldrich, Darmstadt, Germany]. Briefly, tissue microarrays sections (4 μm) were deparaffinized and rehydrated and immersed in the picrosirius red solution for 60 minutes. Physical removal of excess stain was carried out by vigorous slide shaking followed by gentle blotting against soft filter paper. Slides were washed in two changes of acidified water, dehydrate in three changes of 100% ethanol, cleared in xylene, mounted and cover-slipped. Using light microscopy, epithelial cells were stained yellow while the collagen fibres stained red.

Image acquisition

The stained tissue microarrays were scanned using a high-resolution digital scanner (NanoZoomer; Hamamatsu Photonics, Welwyn Garden City, UK), at 20× magnification. For obtaining the images for subsequent analysis, individual tissue spots were exported as JPEG (8-bit) files at ×20 magnification (1368 × 768 pixels). Multiple spots were exported for each core to represent the whole tissue. To improve the reliability of the study, cores with less than 15% area of stroma were excluded from the analysis. All excluded cases were unbiased with respect to clinicopathological features.

Image analysis

The image analysis was carried out using the Fiji package of ImageJ (NIHR public domain) which is available for free download at (https://imagej.nih.gov/ij/download.html). Individual tissue spot JPEG files were imported into ImageJ and was split into RBG (Red, Blue, Green) channels. The image segmentation of tumor from stroma was carried out on the green color space; the red and blue channels were discarded. The segmentation to identify objects of interest was carried out by adjusting the greyscale data to a binary image with foreground pixels of interest set to red (pixel signal intensity of 0) and background pixels set to white (maximum 8-bit intensity of 255). The pixel intensity threshold for classifying all pixels as either foreground or background was defined automatically using the ImageJ histogram thresholding feature. To decease background noise, all images were adjusted to threshold of 150. Fields of interest (specialised stromal collagen surrounding DCIS ducts) (23) were identified by manually delineating of multiple squares, 400x400 μm each in dimension (i.e. equivalent to 2-high power fields of the surrounding collagen), with their bases touching and parallel/tangential to the duct boundary at the area of contact. Several non-overlapping areas were marked per image to segment the surrounding stroma resulting in >100 fields (average 120 fields) of interest were delineated per each core. Features were extracted by measuring the pixel numbers for any particles in the final image using the automated tools in ImageJ. Features extracted included the directionality of collagen fibres, i.e. the angle of the adjacent collagen fibres in each square relative to the angle of the tangent (duct boundary); collagen fibres dispersion, i.e. the arrangement of fibres

relative to each other; and solidity representing the collagen density. Data were recorded in degree (°) for directionality and dispersion while pixels were used for solidity measurement. Each field of interest was analyzed separately and the average reading for each tissue core was considered as the final score for analysis. Moreover, cases with multiple cores were scored and the average of the final scores for each core were used for the analysis. Figure 1 and Supplementary Figures 1 show a schematic representation of the image analysis process.

Calculation of collagen prognostic index score

Each parameter was categorized into three scores (1, 2 and 3) using cut-offs generated from X-tile bioinformatics software (Yale University, version 3.6.1) (24) based local recurrence free interval in the pure DCIS cohort (Supplementary table 1). Score 3 directionality indicated more perpendicular collagen fibres to duct boundary, Score 3 solidity related to dense collagen fibres while Score 3 dispersion referred to lower degree of collagen fibre dispersion i.e. collagen fibres parallel to each other. Collagen prognostic index was calculated according to the equation;

Collagen prognostic index=collagen directionality score (1-3) + collagen solidity score (1-3) + collagen dispersion score (1-3)

The generated score range was 3-9, and was divided into three prognostic groups based on X-tile software whereby a score of 3 or 4 is the good prognostic group, 5-7 for intermediate and score 8 or 9 indicates the poor prognostic group (Supplementary Table 1)

Statistical analysis

Statistical analyses were performed using SPSS v21 (Chicago, IL, USA) for Windows. Spearman's Rho test was used to correlate between the different assessed parameters. Association between collagen prognostic index and its different parameters and clinicopathological parameters was performed using Chi-square, Mann Whitney and Kruskal-Wallis tests (for categorical and numerical data respectively). Univariate survival

analysis against local recurrence free interval was carried out using log rank test and Kaplan Meier curves. Cox regression model was used for multivariate analysis and to evaluate the interaction between radiotherapy and collagen prognostic index regarding the recurrence. For all tests, a two-tailed *p*-value of less than 0.05 was considered as statistically significant.

This work obtained ethics approval by the North West – Greater Manchester Central Research Ethics Committee under the title; Nottingham Health Science Biobank (NHSB), reference number 15/NW/0685.

RESULTS

Distribution of collagen prognostic index parameters and groups

The final number of cases suitable for scoring was 424 pure DCIS and 149 DCIS co-existing with invasive breast cancer after unbiased exclusion of uninformative cores (lost, folded and those containing <15% stroma area). The distribution of collagen prognostic index and the three collagen geometric parameters are summarized in Supplementary Table 2. The median directionality and dispersion angles in the pure DCIS were 19.6^{o} (range 1-85) and 16.7^{o} (range 5-58), respectively, while the median solidity was 0.5 (range 0.001-0.9) and the median collagen prognostic index score was 6 (range 3-9). The distribution of all parameters was unimodal. The poor collagen prognostic index group represented 24% of the cohort with a similar percentage of DCIS in the good prognostic group (23%). Figure 2 shows various patterns of collagen arrangement surrounding DCIS. We found a positive linear correlation between high collagen solidity and perpendicular or oblique collagen arrangement related to the DCIS ducts' boundary (r=0.5, p<0.0001, Spearman's correlation). However, the degree of collagen dispersion showed a negative linear correlation with both collagen solidity (r=-0.3, p<0.0001, Spearman's correlation) and directionality (r=-0.3, p<0.0001, Spearman's correlation).

The proportion of cases with a high collagen prognostic index score (poor prognostic group) was higher in DCIS co-existing with invasive breast cancer than pure DCIS (38% and 24% respectively; χ2=36.5,

p<0.0001) (Table1). Similar results were observed when the score was analysed using a numerical scale (p<0.0001) (Supplementary Table 3). Comparable results were observed for all index parameters whereby DCIS co-existing with invasive carcinoma showed a higher extent of perpendicular collagen fibres in relation to DCIS boundaries and with collagen fibres more often arranged in parallel (lower dispersion) than pure DCIS either in categorical groups or using the numerical scores (Supplementary Tables 4 and 5). The proportion of DCIS-coexisting with invasive breast cancer showed denser collagen fibres was slightly higher than pure DCIS (58% Vs. 50%) which was not statistically significant (p=0.364). Figure 3 illustrates the distribution of collagen prognostic index parameters within pure DCIS and DCIS-invasive breast cancer cohorts.

Significance of collagen prognostic index expression in pure DCIS

Poor collagen prognostic index was associated with other clinicopathological determinants of poor prognosis in DCIS (Table1) including higher nuclear grade (p<0.0001), presence of comedo type necrosis (p<0.0001), hormonal receptor negativity (p=0.032, p=0.029 for ER and PR, respectively) and HER2 positivity (p=0.003). Luminal A DCIS (ER+ and/or PR+, HER2- with <14% proliferation index) had lower collagen prognostic index scores where 67% of cases belonged to good prognostic group (p=0.037). Moreover, poor collagen prognostic index was associated with other microenvironmental and stromal factors associated with poor DCIS prognosis such as dense tumor infiltrating lymphocytes (7) (p=0.016), high levels of prolyl-4-hydroxlases alpha subunit 2 expression (p<0.0001) which is a collagen remodelling protein which was previously reported to be associated with poor outcome in DCIS (10). We validated our results by analyzing the numeric collagen prognostic index scores and found comparable results (Supplementary Table 3). Furthermore, collagen directionality, dispersion and solidity scores were analysed individually with other clinicopathological parameters and the results are summarized in Supplementary Table 4 (for categorical groups) and Supplementary Table 5 (for continuous scores).

Outcome analysis

High collagen prognostic index was associated with shorter DCIS local recurrence free interval either for all recurrences (DCIS or invasive breast cancer) (*HR*=2.0*; 95%CI*=1.3-3.1*; p*=0.002) and for invasive recurrences alone (*HR*=3.1*; 95%CI*=1.6-5.5*; p*=0.0003) (Figure 4). When analysis was confined to patients treated with breast conserving surgery, similar results were obtained where poor collagen prognostic index was associated with worse outcome for all recurrences (*HR*=2.1*; 95%CI*=1.4-3.1*; p*=0.001) and for invasive recurrences (*HR*=3.0*; 95%CI*=1.6-5.4*; p*=0.0003) (Figure 4). Collagen prognostic index was associated with poor outcome regardless of the DCIS nuclear grade, whereby poor collagen prognostic index was associated with shorter local recurrence free interval either for all recurrences or invasive recurrences only (Figure 5) for both high grade DCIS and non-high grade DCIS (low and intermediate DCIS were combined together due to the small number of events in the former group). Moreover, evaluation of the interaction between collagen prognostic index and radiotherapy (collagen prognostic index*radiotherapy) using a Cox regression model showed that higher collagen prognostic index was significantly associated with shorter local recurrence free interval even with radiotherapy (*HR*=1.8, 95%*CI*=1.3-2.3, *p*=0.0001). Analysis of the individual index parameters (collagen directionality, dispersion and solidity), showed significant associations with disease recurrence (Supplementary Figure 2). Forest plots for univariate analysis of other clinicopathological parameters and collagen prognostic index groups and their association with DCIS recurrence are shown in Supplementary Figure 3.

Multivariate survival analysis including the conventional determinants of DCIS risk showed that age at diagnosis, nuclear grade, DCIS size, margin status and comedo type necrosis are independent prognostic factors for DCIS recurrence. However, when collagen prognostic index was incorporated in the model it was the only independent poor prognostic factor for DCIS recurrence in patients treated with breast conserving surgery regardless of other known determinants of high-risk DCIS either for all recurrences (*HR*=2.3, 95% *CI*=1.3-4.2; *p*=0.005) or invasive recurrences only (*HR*=3.4, 95% *CI*=1.5-7.6; *p*=0.003) (Table 2).

DISCUSSION

Precise assessment of DCIS risk is still challenging despite advances in high throughput molecular techniques. None of the currently available clinicopathological based nomograms and indices (25-28) or molecular signatures such as Oncotype DX DCIS (29) provide a reliable and clinically valid tool to predict DCIS risk of progression and/or recurrence to improve personalised management. The biological and clinical heterogeneity of DCIS coupled with underestimation of the important role of the tumor microenvironment in disease behavior are potential reasons for this lack of accuracy. Cancer is not exclusively a disease of malignant cells, but also involves crosstalk between tumor, stromal cells and other components of the surrounding microenvironment. Thus, studying the role of DCIS microenvironment and its interaction with the tumor cells could resolve the DCIS management dilemma and provide an adequate and reliable risk stratification model (1, 5, 30-32).

Extracellular matrix and its main constituent, collagen, plays a key role in normal mammary gland development (33). Interestingly, increased breast tissue density, caused mainly by increased collagen deposition, is a strong risk factors for breast cancer development with high breast density carrying a two- to six-fold increase in breast cancer risk compared to women with normal breast density which is almost as high a risk factor as *BRCA1* and *BRCA2* mutations (4, 34). It is reported that DCIS arises more commonly in dense regions of the breast (4, 35). While was thought that collagen provided a natural physiological barrier against tumor invasion, recent studies have provided concrete evidence on the key role of that collagen dynamic changes (collagen deposition, degradation, re-organization and remodelling) in carcinogenesis, tumor maintenance and progression through its role in cellular adhesion, proliferation, migration, invasion and angiogenesis (36). A recent genetic study showed that extracellular matrix re-organization and response to cellular stress are key factors in breast cancer progression (37). Several studies on invasive breast cancer and other types of cancer using second harmonic generation microscopy or multiphoton laser scanning showed that organization of collagen around malignant tissues provides a tumor associated collagen signature that is associated with tumors behavior and patients' outcome (2, 4, 13). Given that, the role of collagen remodeling in DCIS behavior is relatively understudied, we evaluated the collagen fibre organization surrounding DCIS and its prognostic significance. We have used a novel

imaging method, to the best of our knowledge, to assess collagen characteristics in a large cohort of DCIS and have developed a collagen prognostic index to predict the risk of DCIS recurrence and progression.
In this study, we examined the density and arrangement of collagen relative to each other and the boundary of DCIS. Interestingly, higher collagen prognostic index was associated with features of high-risk DCIS including higher nuclear grade, comedo type necrosis, hormonal receptor negativity and HER2 positivity, supporting its role in DCIS progression. Additionally, our data showed that collagen prognostic index was higher in DCIS co-existing with invasive breast cancer than pure DCIS. Taken together, collagen re-organization is a potential factor for high risk DCIS. The poor prognostic value of collagen rearrangement was shown with a shorter local recurrence free interval in patients with higher levels of collagen prognostic index independently from other clinicopathological factors. These findings were consistent for all recurrent events, either DCIS or invasive breast cancer or when the analysis was confined to invasive recurrences only, which provides more evidence that collagen rearrangement and deposition could play a key role in DCIS progression to invasive disease. collagen prognostic index could predict the risk of recurrence in non-high grade DCIS, which is a group at risk of under-treatment due to the lack of accurate predictors of recurrence in this group, although genetic studies showed that the DCIS are the precursors of low grade invasive breast cancer. Hence, collagen prognostic index could drive the management decision is these patients. The poor prognostic group of non-high grade DCIS had a 22% risk of invasive recurrence within 10 years, compared to zero invasive recurrences in the good prognostic group.

Although the effect of radiotherapy in reduction of DCIS recurrence is undeniable (38-41), there is still a considerable percentage of patients treated with breast conserving surgery and followed by adjuvant radiotherapy that develop disease recurrence. There is no data available for DCIS patients at higher susceptibility for poor response to radiotherapy. Our findings showed that DCIS with a high collagen prognostic index is associated with poor outcome even when radiotherapy was offered. This index could provide a signature for radiotherapy resistance in DCIS. It is reported that dense stroma is associated with

poor response to radiotherapy in cancer treatment (42) which might provide a similar explanation in DCIS. These findings need mechanistic studies for validation.

Although the mechanisms that lead to collagen changes and how these changes affect tumor behavior are not fully understood, several explanations are available. Tumor cells could secret factors that realign a random collagen matrix into a radially arranged collagen to facilitate local invasion (13). The role of hypoxia and collagen remodeling proteins such as prolyl-4-hydroxlases alpha subunit 2 and lysyl oxidase are obvious in developing collagen signatures as shown from previous studies and our findings in the current study (10, 43). As a reciprocal effect, collagen density can affect tumor growth even in tumor harboring scanty stroma. Cells cultured within low-density matrices form well-differentiated acini structures, while colonies that formed within high-density matrices are larger, more proliferative and less organized structures indicating that increasing collagen matrix density can directly promote epithelial cell proliferation (4).

Stiffer extracellular matrix due to increased collagen deposition results in high local mechanical loads and higher resistance to cellular contractility (4). Such changes in the physical microenvironment lead to loss of cellular adhesion, increase tumor cells proliferation and promotes a transformed phenotype. These findings are supported from our results where higher collagen solidity and lower collagen dispersion were associated with a higher proliferation index. Moreover, stromal fibroblasts can regulate epithelial cells in part through secretion of specific soluble growth factors and chemokines such as tumor growth factor β that has been associated with reactive stroma and fibrosis, while studies indicate that the epidermal growth factors/receptors (i.e.HER2) are important not only in the normal mammary gland but also during tumorigenesis and metastasis (44, 45). These findings are supported from our results where higher collagen prognostic index was associated with HER2 overexpression and higher proliferation index.

Mechanical forces can induce epithelial mesenchymal transition leading to tumor invasion (2). With the tumor mass increasing, stress on extracellular matrix increases correspondingly until reaching a critical point, the biomechanical trigger, which can be sensed by mechanical receptors in both tumor and stromal cells. In turn these cells exert cytoskeletal-dependent traction forces on the extracellular matrix (13, 46, 47). Eventually, tumor and stromal cells deform as consequences of the altered tissue tension (48, 49), the

expanding tumor mass (50), matrix stiffening (51), and increased interstitial pressure due to a leaky vasculature and poor lymphatic drainage, initiating epithelial mesenchymal transition (52). These deformed cells acquire a more spindle-like fibroblastic morphology, lower adhesive properties, enhanced motility and invasive behavior. Studies showed that local invasion in stroma rich areas is due an intrinsic property of tumor cells arising in a collagen-dense stroma rather than being a simple reflection of increased local collagen that facilitates invasion. Tumor cells isolated from collagen-dense tissues were in fact innately more migratory, indicating that the ease of invasiveness facilitated by certain collagen structures is likely not the sole cause for increased metastasis but that the tumor cells themselves are more invasive (4). Moreover, aligned collagen provides an adhesive scaffold that support tumor cells in their invasion and migratory process (6, 11-13, 16). Mechanical forces also activate sites on collagen to recruit monocytes leading to a cascade of innate immuno-inflammatory reactions (2). Our results showed that dense collagen is associated with denser tumor infiltrating lymphocytes than lesions with scanty and randomly distributed stromal collagen. The underlying biology warrants an investigation, given our previous report that dense tumor infiltrating lymphocytes have poor prognostic significance in DCIS, a reverse phenomenon to that in invasive disease for which the underlying mechanisms are unclear (7).

This study has been carried out on tissue microarray without including full face sections, which might underestimate the role of DCIS heterogeneity. However, all cases in our cohort were histologically reviewed before tissue microarray construction and multiple cores were used for cases with heterogeneous grade or morphological patterns. Also, our cohort did not include patients treated with endocrine therapy

In conclusion, collagen prognostic index could provide a signature for prediction of DCIS recurrence risk. Collagen deposition and re-organization around DCIS plays a role in tumor progression and recurrence. Collagen features are stable and robust, persisting even when tissues have not been properly fixed, and these features are maintained under various tissue-processing approaches. Thus, picrosirius staining of collagen and subsequent image analysis to assess collagen prognostic index could readily be used in clinical

practice and could drive DCIS management decisions. Additional functional studies to delineate the mechanistic role of collagen re-organization in DCIS behavior are warranted.

Acknowledgment:

This research was supported and funded by the Egyptian Ministry of Higher Education and Scientific Research. KLG was supported by a Yamagiwa-Yoshida Memorial international study grant from the Union for International Cancer Control. We thank the Nottingham Health Science Biobank and Breast Cancer Now Tissue Bank for the provision of tissue samples

References

1. Allen MD, Thomas GJ, Clark S, Dawoud MM, Vallath S, Payne SJ, *et al.* Altered microenvironment promotes progression of preinvasive breast cancer: myoepithelial expression of alphavbeta6 integrin in DCIS identifies high-risk patients and predicts recurrence. Clin Cancer Res. 2014;20:344-57.

2. Fang M, Yuan J, Peng C, Li Y. Collagen as a double-edged sword in tumor progression. Tumor Biol 2014;35:2871-82.

3. Knudsen ES, Ertel A, Davicioni E, Kline J, Schwartz GF, Witkiewicz AK. Progression of ductal carcinoma in situ to invasive breast cancer is associated with gene expression programs of EMT and myoepithelia. Breast Cancer Res Treat. 2012;133:1009-24.

4. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, Rueden CT, *et al.* Collagen density promotes mammary tumor initiation and progression. BMC Med. 2008;6:11.

5. Thompson E, Taube JM, Elwood H, Sharma R, Meeker A, Warzecha HN, *et al.* The immune microenvironment of breast ductal carcinoma in situ. Mod Pathol 2016;29:249-58.

6. Conklin MW, Gangnon RE, Sprague BL, Van Gemert L, Hampton JM, Eliceiri KW, *et al.* Collagen Alignment as a Predictor of Recurrence after Ductal Carcinoma In Situ. Cancer Epidemiol Biomarkers Prev. 2018;27:138-45.

7. Toss MS, Miligy I, Al-Kawaz A, Alsleem M, Khout H, Rida PC, *et al.* Prognostic significance of tumor-infiltrating lymphocytes in ductal carcinoma in situ of the breast. Mod Pathol 2018;31:1226-36.

8. Afghahi A, Forgo E, Mitani AA, Desai M, Varma S, Seto T, *et al.* Chromosomal copy number alterations for associations of ductal carcinoma in situ with invasive breast cancer. Breast Cancer Res. 2015;17:108.

9. Lee S, Stewart S, Nagtegaal I, Luo J, Wu Y, Colditz G, *et al.* Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. Cancer Res 2012;72:4574-86.

10. Toss MS, Miligy IM, Gorringe KL, AlKawaz A, Khout H, Ellis IO, *et al.* Prolyl-4-hydroxylase Alpha subunit 2 (P4HA2) expression is a predictor of poor outcome in breast ductal carcinoma in situ (DCIS). Br J Cancer 2018;119:1518-26.

11. Conklin MW, Eickhoff JC, Riching KM, Pehlke CA, Eliceiri KW, Provenzano PP, *et al.* Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am J Pathol 2011;178:1221-32.

12. Drifka CR, Tod J, Loeffler AG, Liu Y, Thomas GJ, Eliceiri KW, *et al.* Periductal stromal collagen topology of pancreatic ductal adenocarcinoma differs from that of normal and chronic pancreatitis. Mod Pathol 2015;28:1470-80.

13. Provenzano PP, Eliceiri KW, Campbell JM, Inman DR, White JG, Keely PJ. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med 2006;4:38.

14. Vogel B, Siebert H, Hofmann U, Frantz S. Determination of collagen content within picrosirius red stained paraffin-embedded tissue sections using fluorescence microscopy. MethodsX 2015;2:124-34.

15. Whittaker P, Kloner RA, Boughner DR, Pickering JG. Quantitative assessment of myocardial collagen with picrosirius red staining and circularly polarized light. Basic Res Cardiol. 1994;89:397-410.

16. Liu Z, Speroni L, Quinn KP, Alonzo C, Pouli D, Zhang Y, *et al.* 3D organizational mapping of collagen fibers elucidates matrix remodeling in a hormone-sensitive 3D breast tissue model. Biomaterials 2018;179:96-108.

17. Miligy IM, Gorringe KL, Toss MS, Al-Kawaz AA, Simpson P, Diez-Rodriguez M, *et al.* Thioredoxin-interacting protein is an independent risk stratifier for breast ductal carcinoma in situ. Mod Pathol 2018;31:1807-15.

18. Hammond ME, Hayes DF, Wolff AC, Mangu PB, Temin S. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Oncol Pract. 2010;6:195-7.

19. Rakha EA, Pinder SE, Bartlett JM, Ibrahim M, Starczynski J, Carder PJ, *et al.* Updated UK Recommendations for HER2 assessment in breast cancer. J Clin Pathol. 2015;68:93-9.

20. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol. 2011;22:1736-47.

21. van der Groep P, van Diest PJ, Smolders YH, Ausems MG, van der Luijt RB, Menko FH, *et al.* HIF-1alpha overexpression in ductal carcinoma in situ of the breast in BRCA1 and BRCA2 mutation carriers. PloS one 2013;8:e56055.

22. Toss MS, Miligy IM, Gorringe KL, McCaffrey L, AlKawaz A, Abidi A, *et al.* Legumain is an independent predictor for invasive recurrence in breast ductal carcinoma in situ. Mod Pathol 2018; 32:639-649

23. Hendry S, Pang JB, Byrne DJ, Lakhani SR, Cummings MC, Campbell IG, *et al.* Relationship of the Breast Ductal Carcinoma In Situ Immune Microenvironment with Clinicopathological and Genetic Features. Clin Cancer Res 2017;23:5210-7.

24. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004;10:7252-9.

25. Silverstein MJ, Lagios MD, Craig PH, Waisman JR, Lewinsky BS, Colburn WJ, *et al.* A prognostic index for ductal carcinoma in situ of the breast. Cancer 1996;77:2267-74.

26. Lee SK, Yang JH, Woo SY, Lee JE, Nam SJ. Nomogram for predicting invasion in patients with a preoperative diagnosis of ductal carcinoma in situ of the breast. Br J Surg 2013;100:1756-63.

27. Mazouni C, Delaloge S, Rimareix F, Garbay JR. Nomogram for risk of relapse after breast-conserving surgery in ductal carcinoma in situ. J Clin Oncol 2011;29:e44; author reply e5-6.

28. Rudloff U, Jacks LM, Goldberg JI, Wynveen CA, Brogi E, Patil S, *et al.* Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ. J Clin Oncol 2010;28:3762-9.

29. Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C, *et al.* A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. J Natl Cancer Inst. 2013;105:701-10.

30. Semenza GL. The hypoxic tumor microenvironment: A driving force for breast cancer progression. Biochim Biophys Acta. 2016;1863:382-91.

31. Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H, *et al.* Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell 2004;6:17-32.

32. Yeong J, Thike AA, Tan PH, Iqbal J. Identifying progression predictors of breast ductal carcinoma in situ. J Clin Pathol 2017;70:102-8.

33. Hu G, Li L, Xu W. Extracellular matrix in mammary gland development and breast cancer progression. Frontiers Lab Medici 2017;1:36-9.

34. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer 1976;37:2486-92.

35. MacKenzie TA, Titus-Ernstoff L, Vacek PM, Geller B, Weiss JE, Goodrich ME, *et al.* Breast density in relation to risk of ductal carcinoma in situ of the breast in women undergoing screening mammography. Cancer Causes Control 2007;18:939-45.

36. Maragoudakis ME, Missirlis E, Karakiulakis GD, Sarmonica M, Bastakis M, Tsopanoglou N. Basement membrane biosynthesis as a target for developing inhibitors of angiogenesis with anti-tumor properties. Kidney Int 1993;43:147-50.

37. Kothari C, Ouellette G, Labrie Y, Jacob S, Diorio C, Durocher F. Identification of a gene signature for different stages of breast cancer development that could be used for early diagnosis and specific therapy. Oncotarget 2018;9:37407-20.

38. Ben-David MA, Sturtz DE, Griffith KA, Douglas KR, Hayman JA, Lichter AS, *et al.* Long-term results of conservative surgery and radiotherapy for ductal carcinoma in situ using lung density correction: the University of Michigan experience. Breast J 2007;13:392-400.

39. Bijker N, Meijnen P, Peterse JL, Bogaerts J, Van Hoorebeeck I, Julien JP, *et al.* Breast-conserving treatment with or without radiotherapy in ductal carcinoma-in-situ: ten-year results of European Organisation for Research and Treatment of Cancer randomized phase III trial 10853--a study by the EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group. J Clin Oncol 2006;24:3381-7.

40. Dodwell D, Clements K, Lawrence G, Kearins O, Thomson CS, Dewar J, *et al.* Radiotherapy following breast-conserving surgery for screen-detected ductal carcinoma in situ: indications and utilisation in the UK. Interim findings from the Sloane Project. Br J Cancer 2007;97:725-9.

41. Donker M, Litiere S, Werutsky G, Julien JP, Fentiman IS, Agresti R, *et al.* Breast-conserving treatment with or without radiotherapy in ductal carcinoma In Situ: 15-year recurrence rates and outcome after a recurrence, from the EORTC 10853 randomized phase III trial. J Clin Oncol 2013;31:4054-9.

42. Barker HE, Paget JTE, Khan AA, Harrington KJ. The tumor microenvironment after radiotherapy: mechanisms of resistance and recurrence. Nat Rev Cancer 2015;15:409-25.

43. Gilkes DM, Chaturvedi P, Bajpai S, Wong CC, Wei H, Pitcairn S, *et al.* Collagen prolyl hydroxylases are essential for breast cancer metastasis. Cancer Res 2013;73:3285-96.

44. Orimo A, Gupta PB, Sgroi DC, Arenzana-Seisdedos F, Delaunay T, Naeem R, *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. Cell 2005;121:335-48.

45. Bavik C, Coleman I, Dean JP, Knudsen B, Plymate S, Nelson PS. The gene expression program of prostate fibroblast senescence modulates neoplastic epithelial cell proliferation through paracrine mechanisms. Cancer Res 2006;66:794-802.

46. Roth JM, Caunt M, Cretu A, Akalu A, Policarpio D, Li X, *et al.* Inhibition of experimental metastasis by targeting the HUIV26 cryptic epitope in collagen. Am J Pathol 2006;168:1576-86.

47. Meyaard L. The inhibitory collagen receptor LAIR-1 (CD305). J Leukoc Biol. 2008;83:799-803.

48. Provenzano PP, Alejandro-Osorio AL, Valhmu WB, Jensen KT, Vanderby R, Jr. Intrinsic fibroblast-mediated remodeling of damaged collagenous matrices in vivo. Matrix Biol 2005;23:543-55.

49. Suresh S. Biomechanics and biophysics of cancer cells. Acta Biomater 2007;3:413-38.

50. Paszek MJ, Weaver VM. The tension mounts: mechanics meets morphogenesis and malignancy. J Mammary Gland Biol Neoplasia 2004;9:325-42.

51. Paszek MJ, Zahir N, Johnson KR, Lakins JN, Rozenberg GI, Gefen A, *et al.* Tensional homeostasis and the malignant phenotype. Cancer Cell 2005;8:241-54.

52. Netti PA, Berk DA, Swartz MA, Grodzinsky AJ, Jain RK. Role of extracellular matrix assembly in interstitial transport in solid tumors. Cancer Res 2000;60:2497-503.

Figures legends:

Figure 1: Diagram shows the different parameters assessed to evaluate the collagen characteristics surrounding DCIS. Green squares refer to the fields of interest for analysis (400x400 µm surrounding the DCIS boundary and note that all squares are aligned touching and parallel to imaginary tangential line to the duct boundary. A) Perpendicularly arranged collagen fibres relative to duct boundary, B) Parallel arranged fibres towards the duct boundary, C) Collagen fibres arranged parallel to each other (lower dispersion) while D) randomly arranged fibres (higher dispersion). E) Dense collagen fibres and F) scanty collagen fibres. During the analysis the 3 parameters were assessed per each field of interest (square) and the final average value was used in the analysis.

Figure 2: Example of collagen fibre arrangement in 2 DCIS cases. A and B represent a case surrounded by dense collagen with lower dispersion (parallelly arranged collagen fibres all over the duct perimetry) and different directionalities, where collagen fibres delineated in (B) with black square show parallel arrangement relative to duct boundary (The tangential line), the fibres in the green square show more perpendicular arrangement relative to the tangential line touching the duct boundary. C and D represent DCIS surrounded by scanty randomly distributed collagen fibres with varied directionality relative to tangential line to the duct boundary. (A and C are DCIS sections stained with picrosirius red, while B and D are the images generated from ImageJ software).

Figure 3: Violin plots comparing the distribution of collagen prognostic index parameters between pure DCIS and DCIS co-existing with invasive carcinoma (mixed DCIS). Obviously, DCIS mixed showed higher angle of directionality and lower dispersion degree than pure DCIS. No statistically significant difference between collagen density between pure and mixed DCIS. Central line in the boxplot represents the median while the error bars represent the standard deviation.

Figure 4: Kaplan Meier curves show that the poor collagen prognostic index group is associated with shorter ipsilateral local recurrence free interval for all recurrences (A) and invasive recurrences (B) in the whole series as well as in the breast conserving surgery treated group (C; for all recurrences and D; for invasive recurrences).

Figure 5: Kaplan Meier curves show that poor collagen prognostic index group is associated with shorter ipsilateral local recurrence free interval for all recurrences (A) and invasive recurrences (B) in non-high grade DCIS as well as in high grade DCIS (C; for all recurrences and D; for invasive recurrences).

Table 1: Correlation between collagen prognostic index and clinicopathological parameters in the pure DCIS cohort

Clinicopathological Parameters	Collagen prognostic index groups			χ^2 (*p*-value)
	Good (n=96)	Moderate (n=128)	Poor (n=100)	
Age (years)				
≤50	26 (27)	52 (23)	30 (30)	2.1
>50	70 (73)	176 (77)	70 (70)	(0.356)
Presentation				
Screening	45 (47)	115 (50)	60 (60)	3.7
Symptomatic	51 (53)	113 (50)	40 (40)	(0.150)
DCIS Size (mm)				
≤20	45 (47)	90 (40)	45 (45)	1.7
>20	51 (53)	137 (60)	55 (55)	(0.414)
Nuclear Grade				
Low	39 (41)	7 (3)	0 (0)	155.6
Moderate	35 (37)	64 (28)	9 (9)	**(<0.0001)**
High	22 (22)	157 (69)	91 (91)	
Comedo necrosis				
Yes	41 (43)	168 (74)	83 (83)	42.4
No	55 (57)	60 (26)	17 (17)	**(<0.0001)**
Estrogen receptor status				
Negative	15 (17)	59 (28)	32 (34)	6.8
Positive	73 (83)	154 (72)	62 (66)	**(0.032)**
Progesterone receptor				
Negative	28 (32)	99 (46)	47 (49)	7.1
Positive	61 (69)	115 (54)	49 (51)	**(0.029)**
HER2 status				
Negative	79 (86)	163 (75)	63 (64)	11.8
Positive	13 (14)	54 (25)	35 (36)	**(0.003)**
Proliferation index (Ki-67)				
High	13 (16)	49 (24)	29 (31)	5.4
Low	68 (84)	156 (76)	64 (69)	(0.066)
Molecular classes				
Luminal A	51 (67)	95 (49)	34 (40)	13.4
Luminal B	11 (14)	40 (21)	20 (24)	**(0.037)**
HER2 Enriched	6 (8)	32 (17)	19 (22)	
Triple negative	8 (11)	25 (13)	12 (14)	
Tumor infiltrating lymphocytes				
Dense	27 (39)	96 (50)	53 (62)	8.2
Sparse	43 (61)	96 (50)	33 (38)	**(0.016)**
P4HA2 expression				
High	28 (34)	127 (62)	52 (60)	19.2
Low	54 (66)	78 (38)	35 (40)	**(<0.0001)**
HIF1α expression				
High	11 (15)	45 (25)	23 (30)	4.7
Low	62 (85)	137 (75)	54 (70)	0.094
Tumor Type*				
Pure DCIS	96 (97)	228 (72)	100 (64)	36.5
DCIS associated with IBC	3 (3)	89 (28)	57 (36)	**(<0.0001)**

*Including pure DCIS (n=424) and DCIS coexisting with IBC (n=149)
Significant p values are in **bold**
DCIS; ductal carcinoma in situ, HER2; human epidermal growth factor receptor 2, P4HA2; prolyl-4-hydroxylase alpha subunit 2, HIF1α; hypoxia inducible factor-1 alpha, IBC; invasive breast cancer.

Table 2: Multivariate survival analysis (Cox regression model) of variables predicting outcome in terms of ipsilateral local recurrence (either all recurrences A, or invasive recurrences B) in DCIS patients treated by breast conserving surgery.

A) All recurrences

Conventional clinicopathological parameters associated with high risk DCIS				
Parameters	Hazard ratio (*HR*)	95% *confidence interval (CI)*		*p*-value
		Lower	Upper	
Patient Age	0.4	0.2	0.8	**0.006**
DCIS presentation	1.5	0.9	2.4	0.111
DCIS size	1.5	1.1	2.1	**0.040**
DCIS nuclear Grade	1.9	1.3	2.7	**0.001**
Comedo necrosis	0.6	0.4	0.9	**0.049**
Margin status	0.8	0.7	0.9	**0.004**
Collagen prognostic index and other clinicopathological parameters associated with high risk DCIS				
Collagen prognostic index	2.3	1.3	4.2	**0.005**
Patient Age	0.5	0.2	1.1	0.072
DCIS presentation	2.1	0.9	4.3	0.052
DCIS size	0.9	0.5	1.6	0.760
DCIS nuclear Grade	1.1	0.5	1.9	0.994
Comedo necrosis	0.8	0.4	1.6	0.559
Margin status	0.9	0.8	1.1	0.236

Significant *p* values are in **bold**

B) Invasive recurrence

Conventional clinicopathological parameters associated with high risk DCIS				
Parameters	Hazard ratio (*HR*)	95% *confidence interval (CI)*		*p*-value
		Lower	Upper	
Patient Age	0.5	0.2	1.3	0.156
DCIS presentation	1.4	0.8	2.6	0.245
DCIS size	1.8	1.1	2.4	**0.013**
DCIS nuclear Grade	1.9	1.1	3.0	**0.013**
Comedo necrosis	0.7	0.4	1.3	0.274
Margin status	0.9	0.8	1.1	0.075
Collagen prognostic index and other clinicopathological parameters associated with high risk DCIS				
Collagen prognostic index	3.4	1.5	7.6	**0.003**
Patient Age	1.2	0.4	3.7	0.697
DCIS presentation	3.0	1.2	7.9	**0.025**
DCIS size	1.3	0.7	2.7	0.412
DCIS nuclear Grade	0.9	0.4	2.3	0.909
Comedo necrosis	0.9	0.4	2.1	0.786
Margin status	0.9	0.8	1.1	0.838

Significant *p* values are in **bold**

Figure 1

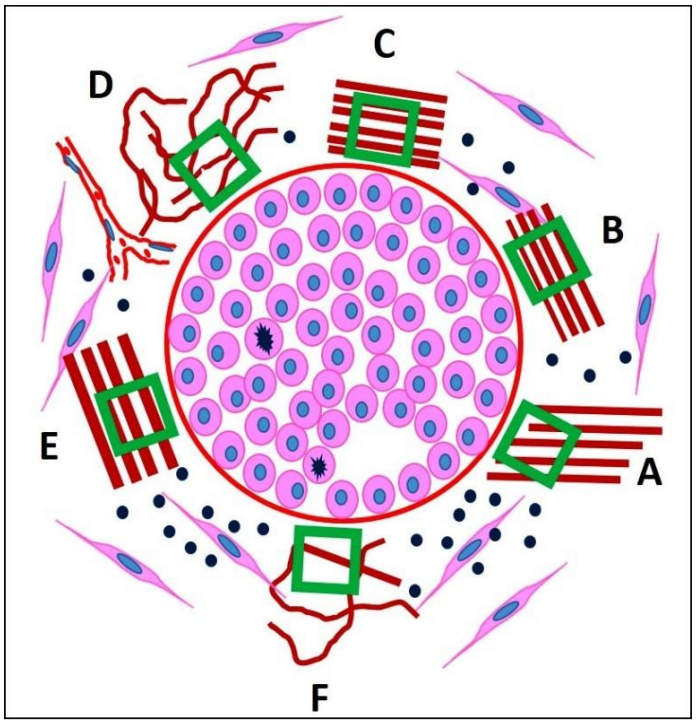


Figure 2

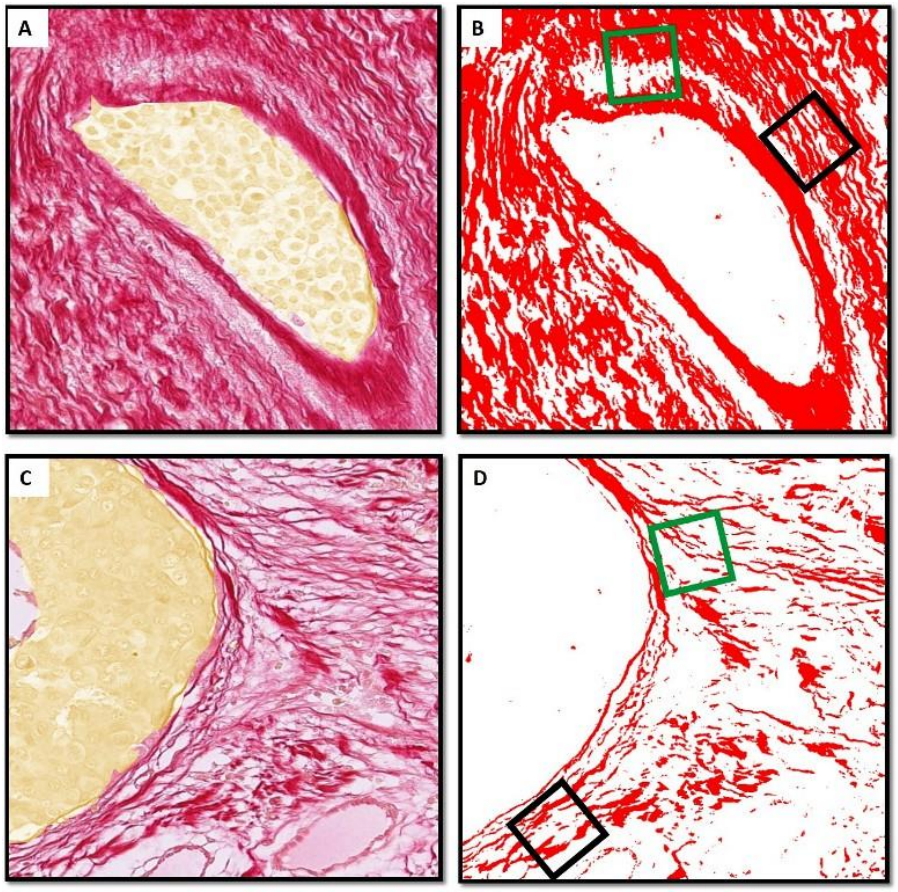


Figure 3

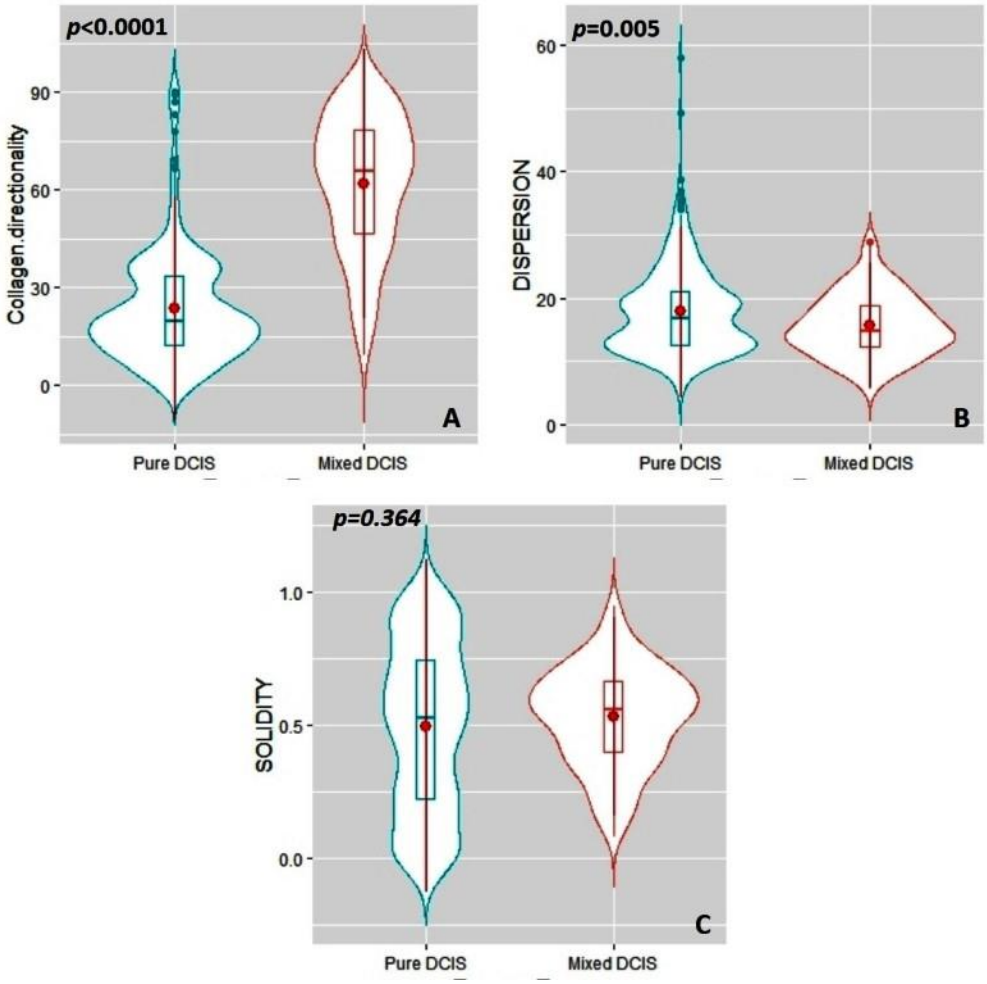

p<0.0001
Collagen.directionality
90
60
30
0
A
Pure DCIS
Mixed DCIS
p=0.005
DISPERSION
60
40
20
0
B
Pure DCIS
Mixed DCIS
p=0.364
SOLIDITY
1.0
0.5
0.0
C
Pure DCIS
Mixed DCIS

Figure 4

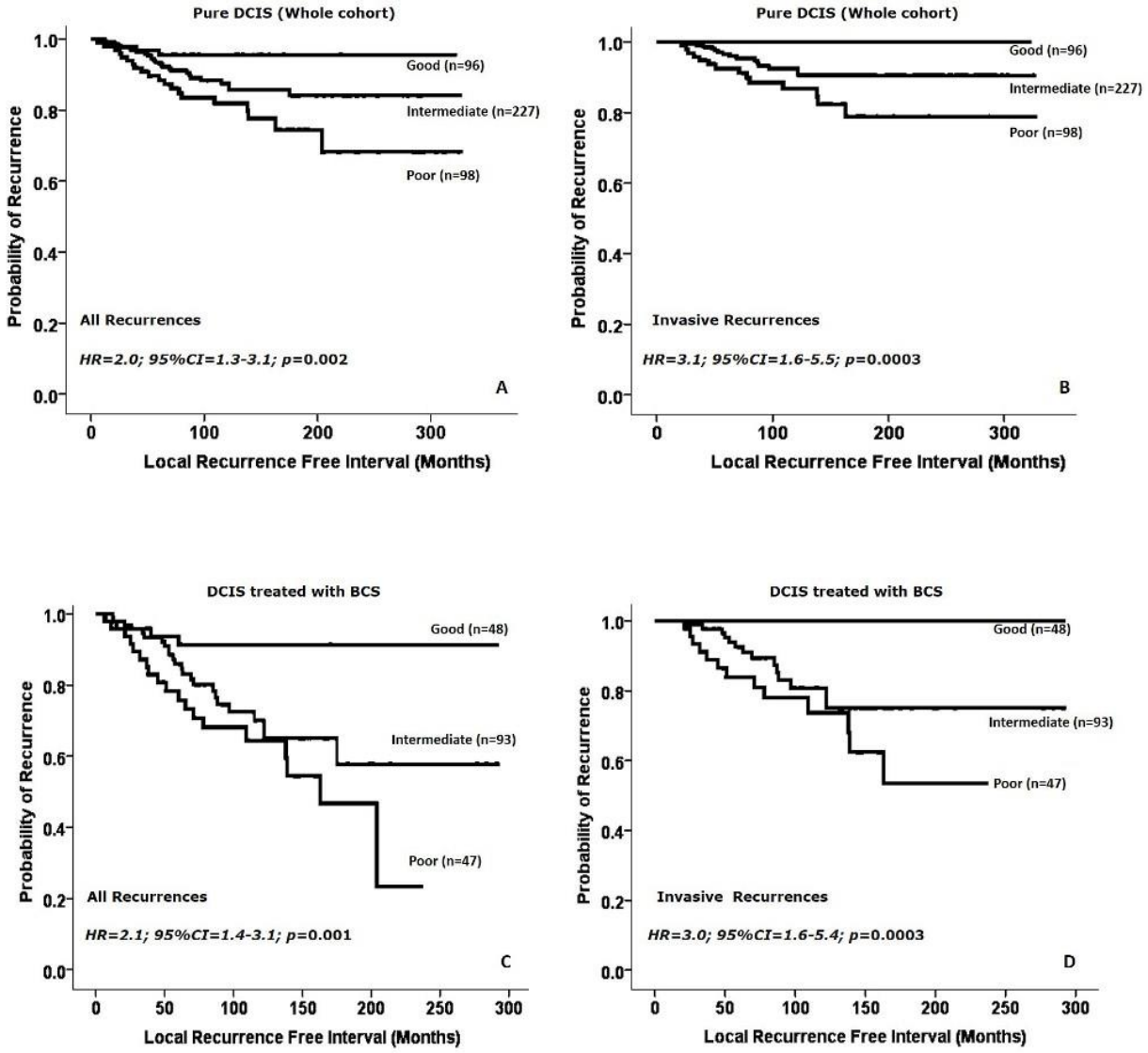
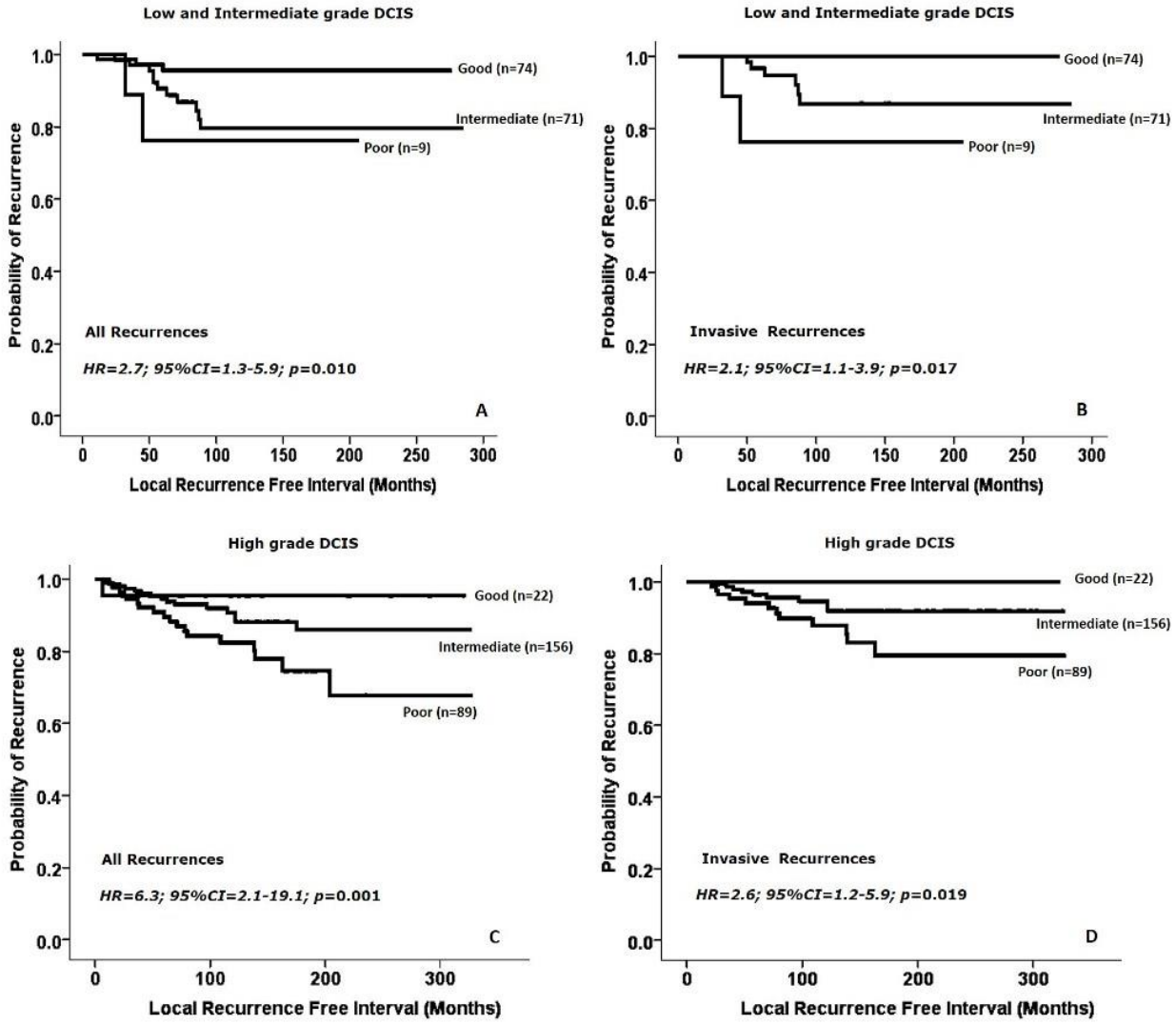

Pure DCIS (Whole cohort)
Probability of Recurrence
1.0
0.8
0.6
0.4
0.2
0.0
Good (n=96)
Intermediate (n=227)
Poor (n=98)
All Recurrences
HR=2.0; 95%CI=1.3-3.1; p=0.002
A
0
100
200
300
Local Recurrence Free Interval (Months)
Pure DCIS (Whole cohort)
Probability of Recurrence
Good (n=96)
Intermediate (n=227)
Poor (n=98)
Invasive Recurrences
HR=3.1; 95%CI=1.6-5.5; p=0.0003
B
Local Recurrence Free Interval (Months)
DCIS treated with BCS
Probability of Recurrence
Good (n=48)
Intermediate (n=93)
Poor (n=47)
All Recurrences
HR=2.1; 95%CI=1.4-3.1; p=0.001
C
0
50
100
150
200
250
300
Local Recurrence Free Interval (Months)
DCIS treated with BCS
Probability of Recurrence
Good (n=48)
Intermediate (n=93)
Poor (n=47)
Invasive Recurrences
HR=3.0; 95%CI=1.6-5.4; p=0.0003
D
Local Recurrence Free Interval (Months)

Figure 5


Low and Intermediate grade DCIS
Probability of Recurrence
Local Recurrence Free Interval (Months)
Good (n=74)
Intermediate (n=71)
Poor (n=9)
All Recurrences
HR=2.7; 95%CI=1.3-5.9; p=0.010
A
Low and Intermediate grade DCIS
Probability of Recurrence
Local Recurrence Free Interval (Months)
Good (n=74)
Intermediate (n=71)
Poor (n=9)
Invasive Recurrences
HR=2.1; 95%CI=1.1-3.9; p=0.017
B
High grade DCIS
Probability of Recurrence
Local Recurrence Free Interval (Months)
Good (n=22)
Intermediate (n=156)
Poor (n=89)
All Recurrences
HR=6.3; 95%CI=2.1-19.1; p=0.001
C
High grade DCIS
Probability of Recurrence
Local Recurrence Free Interval (Months)
Good (n=22)
Intermediate (n=156)
Poor (n=89)
Invasive Recurrences
HR=2.6; 95%CI=1.2-5.9; p=0.019
D

Supplementary Table 1: Scores of collagen directionality, dispersion, density and collagen prognostic index based on cut-off points generated from X-tile software

	Collagen Directionality			Collagen Dispersion			Collagen Solidity		
Score	1	2	3	1	2	3	1	2	3
Threshold	<13.6°	13.6-19.6°	>19.6°	>15.8°	12.3-15.8°	<12.3°	<0.2	0.2-0.6	>0.6

Collagen prognostic index (CPI) = Collagen directionality score + collagen solidity score + collagen dispersion score

Collagen prognostic index groups:

- Good prognostic group: Scores 3 or 4
- Moderate prognostic group: Scores 5, 6 or 7
- Poor prognostic group: Scores 8 or 9

Supplementary Table 2: Distribution of collagen prognostic index and its parameters scores within the pure DCIS cohort

	Collagen Directionality	Collagen Dispersion	Collagen Solidity	CPI
Mean ± SD	23.6 ± 17.2	17.8 ± 6.8	0.5 ± 0.3	6 ± 1.8
Median	19.6	16.7	0.5	6
Minimum	1.0	5.1	0.001	3
Maximum	85.0	58.1	0.9	9

Supplementary Table 3: Correlation between collagen prognostic index scores (continuous values) and different clinicopathological parameters in the pure DCIS cohort.

Clinicopathological Parameters	Number of cases	collagen prognostic index score	
		Mean Rank	*p*-value
Age (years)			
≤50	108	216.9	0.657
>50	316	210.9	
Presentation			
Screening	220	224.0	**0.042**
Symptomatic	204	200.0	
DCIS Size (mm)			
≤20	180	214.2	0.742
>20	243	210.3	
Nuclear Grade			
Low	46	65.8	**<0.0001**
Moderate	108	155.3	
High	270	260.4	
Comedo necrosis			
Yes	292	237.7	**<0.0001**
No	132	156.9	
Estrogen receptor			
Negative	106	221.2	**0.013**
Positive	289	189.5	
Progesterone Receptor			
Negative	174	215.7	**0.015**
Positive	225	187.8	
HER2 status			
Negative	305	193.9	**0.003**
Positive	102	234.0	
Proliferation index (Ki-67)			
High	91	218.1	**0.004**
Low	288	18611	
Molecular classes			
Luminal A	180	160.4	**0.017**
Luminal B	71	192.4	
HER2 Enriched	57	198.6	
Triple negative	45	191.8	
Tumour infiltrating lymphocytes			
Dense	176	185.1	**0.044**
Sparse	172	163.7	
P4HA2 expression			
High	207	206.4	**0.0001**
Low	167	164.2	
HIF1A expression			
High	79	188.1	**0.017**
Low	253	159.5	
Tumour Type*			
Pure DCIS	424	262.9	**<0.0001**
DCIS associated with IBC	149	355.6	

*Including pure DCIS (n=424) and DCIS coexisting with IBC (n=149)

Significant p values are in **bold**
DCIS; ductal carcinoma in situ, HER2; human epidermal growth factor receptor 2, P4HA2; prolyl-4-hydroxylase alpha subunit 2, HIF1α; hypoxia inducible factor-1 alpha, IBC; invasive breast cancer.

Supplementary Table 4: Correlation between collagen prognostic index parameters and different clinicopathological parameters in the pure DCIS cohort.

Clinicopathological Parameters	Collagen Directionality (n=424)			χ^2 (*p*-value)	Collagen Solidity (n=424)			χ^2 (*p*-value)	Collagen Dispersion (n=424)			χ^2 (*p*-value)
	Score 1	Score 2	Score 3		Score 1	Score 2	Score 3		Score 1	Score 2	Score 3	
Age (years)												
≤50	33 (26)	25 (28)	50 (24)	0.6	25 (26)	43 (28)	40 (23)	0.9	54 (23)	21 (22)	33 (34)	4.9
>50	92 (74)	64 (72)	160 (76)	(0.710)	72 (74)	111 (72)	133 (77)	(0.608)	179 (77)	73 (78)	64 (66)	(0.088)
Presentation												
Screening	60 (48)	37 (42)	123 (59)	8.3	42 (43)	82 (53)	96 (56)	3.8	116 (50)	55 (58)	49 (51)	2.1
Symptomatic	65 (52)	52 (59)	87 (41)	**(0.016)**	55 (57)	72 (47)	77 (44)	(0.144)	117 (50)	39 (42)	48 (49)	(0.343)
DCIS Size (mm)												
≤20	54 (44)	31 (35)	95 (45)	2.8	45 (46)	65 (43)	70 (41)	0.8	95 (41)	44 (47)	41 (43)	0.9
>20	70 (56)	58 (65)	115 (55)	(0.242)	52 (54)	88 (57)	103 (59)	(0.639)	138 (59)	50 (53)	55 (57)	(0.607)
Nuclear Grade												
Low	30 (24)	15 (17)	1 (1)	54.5	46 (47)	0 (0)	0 (0)		32 (14)	11 (12)	3 (3)	47.3
Moderate	36 (29)	16 (18)	56 (27)	**(<0.0001)**	44 (44)	50 (33)	14 (8)	164.2	84 (36)	15 (16)	9 (9)	**(<0.0001)**
High	59 (47)	58 (65)	153 (72)		7 (7)	104 (67)	159 (92)	**(<0.0001)**	117 (50)	68 (72)	85 (88)	
Comedo necrosis												
Yes	73 (58)	62 (70)	157 (75)	9.8	32 (33)	114 (74)	146 (84)	79.6	144 (62)	70 (75)	78 (80)	12.8
No	52 (42)	27 (30)	53 (25)	**(0.007)**	65 (67)	40 (26)	27 (16)	**(<0.0001)**	89 (38)	24 (25)	19 (20)	**(0.002)**
Estrogen receptor status												
Negative	27 (24)	30 (36)	49 (25)	4.3	8 (9)	48 (33)	50 (31)	19.8	48 (22)	23 (26)	35 (38)	7.7
Positive	87 (76)	54 (64)	148 (75)	(0.114)	83 (91)	96 (67)	110 (69)	**(0.0001)**	166 (78)	65 (74)	58 (62)	**(0.022)**
Progesterone receptor												
Negative	48 (41)	47 (56)	79 (40)	6.7	20 (22)	73 (50)	81 (50)	23.3	82 (38)	41 (47)	51 (54)	7.3
Positive	68 (59)	37 (44)	120 (60)	**(0.036)**	72 (78)	72 (50)	81 (50)	**(<0.0001)**	135 (62)	46 (53)	44 (46)	**(0.025)**
HER2 status												
Negative	93 (78)	57 (68)	155 (76)	2.9	88 (95)	107 (72)	110 (66)	26.4	175 (79)	67 (74)	63 (66)	5.9
Positive	27 (22)	27 (32)	48 (24)	**(0.237)**	5 (5)	41 (28)	56 (34)	**(<0.0001)**	46 (21)	24 (26)	32 (34)	(0.051)
Proliferation index (Ki-67)												
High	23 (21)	19 (25)	49 (25)	0.7	11 (13)	35 (25)	45 (29)	7.2	37 (18)	23 (28)	31 (33)	9.1
Low	86 (79)	58 (75)	144 (75)	(0.696)	71 (87)	107 (75)	110 (71)	**(0.027)**	167 (82)	59 (72)	62 (67)	**(0.011)**

Molecular classes												
Luminal A	58 (56)	31 (42)	91 (52)	7.2 (0.303)	59 (76)	61 (47)	60 (41)	32.1 **(<0.0001)**	108 (56)	36 (48)	36 (42)	8.8 (0.185)
Luminal B	19 (18)	15 (20)	37 (21)		10 (13)	22 (17)	39 (26)		37 (19)	17 (23)	17 (20)	
HER2 Enriched	16 (16)	18 (24)	23 (13)		2 (3)	27 (21)	28 (19)		23 (12)	13 (17)	21 (24)	
Triple negative	10 (10)	10 (14)	25 (14)		6 (8)	19 (15)	20 (14)		24 (13)	9 (12)	12 (14)	
Tumour infiltrating lymphocytes												
Dense	49 (50)	41 (53)	86 (50)	0.3 (0.867)	28 (38)	64 (51)	84 (56)	6.8 **(0.033)**	85 (46)	39 (49)	52 (62)	5.8 (0.054)
Sparse	49 (50)	36 (47)	87 (50)		46 (62)	61 (49)	65 (44)		99 (54)	41 (51)	32 (38)	
P4HA2 expression												
High	50 (46)	48 (58)	109 (60)	5.6 (0.060)	27 (32)	89 (63)	91 (62)	24.7 **(<0.0001)**	99 (49)	50 (58)	58 (67)	7.8 **(0.020)**
Low	59 (54)	34 (42)	74 (40)		58 (68)	53 (37)	56 (38)		102 (51)	36 (42)	29 (33)	
HIF1A expression												
High	23 (24)	20 (28)	36 (22)	0.9 (0.608)	9 (12)	26 (21)	44 (34)	14.2 **(0.001)**	38 (20)	16 (22)	25 (34)	5.7 (0.057)
Low	72 (76)	52 (72)	129 (78)		69 (88)	98 (79)	86 (66)		149 (80)	56 (78)	48 (66)	
Tumour Type*												
Pure DCIS	125 (98)	89 (95)	210 (60)	97.8 **(<0.0001)**	97 (91)	154 (65)	173 (75)	26.4 **(<0.0001)**	233 (78)	94 (67)	97 (72)	6.8 **(0.033)**
DCIS associated with IBC	2 (2)	5 (5)	142 (40)		9 (9)	82 (35)	58 (25)		65 (22)	47 (33)	37 (27)	

*Including pure DCIS (n=424) and DCIS coexisting with IBC (n=149)

Significant p values are in **bold**

DCIS; ductal carcinoma in situ, HER2; human epidermal growth factor receptor 2, P4HA2; prolyl-4-hydroxylase alpha subunit 2, HIF1α; hypoxia inducible factor-1 alpha, IBC; invasive breast cancer.

Supplementary Table 5: Correlation between different collagen prognostic index parameters and clinicopathological parameters in the pure DCIS cohort using continuous data.

Clinicopathological Parameters	Number of cases	Collagen Directionality		Collagen Solidity		Collagen Dispersion	
		Mean Rank	*p*-value	Mean Rank	*p*-value	Mean Rank	*p*-value
Age (years)							
≤50	108	210.2	0.822	201.7	0.286	198.3	0.162
>50	316	213.3		216.2		217.4	
Presentation							
Screening	220	225.0	**0.029**	225.2	**0.027**	211.4	0.840
Symptomatic	204	199.0		198.8		213.8	
DCIS Size (mm)							
≤20	180	216.9	0.482	207.2	0.489	213.7	0.801
>20	243	208.4		215.5		210.7	
Nuclear Grade							
Low	46	83.0	**<0.0001**	31.4	**<0.0001**	249.1	**<0.0001**
Moderate	108	192.2		134.2		264.2	
High	270	242.7		274.7		185.6	
Comedo necrosis							
Yes	292	227.0	**0.0002**	244.8	**<0.0001**	202.7	**0.014**
No	132	180.4		140.9		234.2	
Estrogen receptor							
Negative	106	200.7	0.776	230.4	**0.001**	175.7	**0.019**
Positive	289	197.0		186.1		206.2	
Progesterone Receptor							
Negative	174	197.7	0.722	226.3	**<0.0001**	185.7	**0.029**
Positive	225	201.8		179.6		211.1	
HER2 status							
Negative	305	199.5	0.183	189.2	**<0.0001**	210.5	0.055
Positive	102	217.4		248.3		184.6	
Proliferation index (Ki-67)							
High	91	200.6	0.290	211.8	**0.029**	161.7	**0.005**
Low	288	186.7		183.1		198.9	
Molecular classes							
Luminal A	180	171.5	0.694	153.9	**0.0001**	186.6	0.109
Luminal B	71	186.4		198.8		176.8	
HER2 Enriched	57	175.8		208.2		148.6	
Triple negative	45	185.6		195.5		174.7	
Tumour infiltrating lymphocytes							
Dense	176	175.2	0.892	187.3	**0.016**	165.1	0.075
Sparse	172	173.7		163.3		184.2	
P4HA2 expression							
High	207	197.4	**0.048**	205.6	**<0.0001**	175.3	**0.015**
Low	167	175.2		165.1		202.7	
HIF1A expression							
High	79	168.1	0.865	198.9	**0.001**	146.9	**0.038**
Low	253	166.0		156.4		172.6	

Tumour Type*							
Pure DCIS	424	225.8	**<0.0001**	283.3	0.364	298.6	**0.005**
DCIS associated with IBC	149	461.3		297.6		254.0	

*Including pure DCIS (n=424) and DCIS coexisting with IBC (n=149)
Significant p values are in **bold**
DCIS; ductal carcinoma in situ, HER2; human epidermal growth factor receptor 2, P4HA2; prolyl-4-hydroxylase alpha subunit 2, HIF1α; hypoxia inducible factor-1 alpha, IBC; invasive breast cancer.