

Collagen (XI) alpha-1 chain (COL11A1) is an independent prognostic factor in breast ductal carcinoma in situ

Michael S Toss^{1,2}, Islam M Miligy^{1,3}, Kylie L Gorringer^{4,5}, Mohammed A Aleskandarany^{1,3}, Abdulbaqi Alkawaz¹, Karuna Mittal⁶, Ritu Aneja⁶, Ian O Ellis¹, Andrew R Green¹ and Emad A Rakha^{1,3}

¹Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham, Nottingham City Hospital, UK.

²Histopathology department, South Egypt Cancer Institute, Assiut University, Egypt,

³Histopathology department, Faculty of Medicine, Menoufia University, Egypt, ⁴Cancer Genomics Program, Peter MacCallum Cancer Centre, Melbourne, ⁵The Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Australia and

⁶Georgia State University, Atlanta, GA, USA.

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, COL11A1, stromal expression, recurrence, radiotherapy

Running title: COL11A1 expression in DCIS

Abstract

Collagen11A1 (COL11A1) is a fibrillary type collagen constituting a minor component of the extracellular matrix and plays role in tissue tensile strength. Overexpression of COL11A1 expression is associated with aggressive behavior and poor outcome in several human malignancies. In this study, we evaluated the association between COL11A1 expression and clinicopathological parameters of the breast ductal carcinoma *in situ* (DCIS) and its prognostic value. COL11A1 protein expression was assessed immunohistochemically in a large well-characterized cohort of DCIS including pure (n=776) and DCIS associated with invasive carcinoma (DCIS-mixed, n=239). COL11A1 expression was assessed in tumor cells and surrounding stromal cells and correlated with clinicopathological parameters, immunoprofile and disease outcome. In pure DCIS, high COL11A1 expression was observed in tumor cells and surrounding stromal cells in 25% and 13% of cases, respectively. Higher COL11A1 expression within the stromal cells was associated with hormone receptor negative, HER2 enriched and triple negative molecular subtypes and showed a positive linear correlation with proliferation index, dense tumor infiltrating lymphocytes and hypoxia-inducible factor 1 alpha. COL11A1 expression in tumor and stromal cells was significantly higher in DCIS associated with invasive carcinoma than in pure DCIS, and within the DCIS-mixed cohort, the invasive component showed higher COL11A1 expression than the DCIS component (all, $p < 0.0001$). Overexpression of stromal COL11A1 was an independent predictor of shorter local recurrence free interval for all recurrences ($HR = 13.2$, $95\%CI = 6.9-25.4$, $p < 0.0001$) and for invasive recurrences ($HR = 11.2$, $95\%CI = 4.9-25.8$, $p < 0.0001$). When incorporated with other risk factors, stromal COL11A1 provided better patient risk stratification. DCIS with higher stromal COL11A1 expression showed poor outcome even with adjuvant radiotherapy management.

Conclusion: Overexpression of stromal COL11A1 is associated with invasive recurrence in DCIS and is a potential marker to predict the response to radiotherapy.

INTRODUCTION

Invasive breast carcinoma accounts for half of breast ductal carcinoma in situ (DCIS) recurrences (1-6). Several prognostic indices and molecular signatures have been

devised to assess the potential risk in DCIS towards its progression into invasive disease and to anticipate disease recurrence. However, the reliability and reproducibility of these indices are still insufficient for precise management decision making. Lack of reproducibility and reliability of these signatures and indices might be a result of their focus on imprecise clinical parameters only such as Van Nuys Prognostic Index (4), or depending on molecular genetic signatures like Oncotype DX DCIS, which has been updated to incorporate other clinical data such patient age and DCIS size by the Genomic Health (www.genomichealth.com) (2). However, the latter mRNA-based signatures ignore the potential effect of post-transcriptional and post-translational modifications resulting in the differential expression of proteins within the tumor cells and subsequent effects of such changes on their functions. One more limitation of these assays is the collective measurement of gene expression levels in the tested tumor tissue sample, with variable proportions of tumor and stromal elements, therefore, omitting or underestimating the role of tumor microenvironment. Cancer is not exclusively a disease of tumor cells, but a disease of wider derangement and crosstalk involving tumor epithelial cells, surrounding stroma and tumor microenvironment including immune infiltrate. Recently, the role of microenvironment has been magnified in cancer development and progression and considered as a main component of cancer hallmarks (7). A supportive microenvironment could be crucial for remnant DCIS cells to survive therapy and develop invasive capabilities. Therefore, identification of more robust genetic and/or proteomic signatures incorporating the crosstalk between tumor epithelial cells and surrounding microenvironment might provide a better approach for DCIS risk assessment and hence better personalised management to avoid over- and under-treatment.

Extracellular matrix and its main constituent, collagen, plays a critical role in several biological processes in normal tissues. Collagen is aberrantly produced by cancer associated fibroblasts, which reciprocally interact with tumor cells promoting the latter's proliferation, migration and differentiation and, therefore, aiding cancer development and progression (8).

Collagen11A1 (COL11A1) is a type of fibrillary collagen encoded by the *COL11A1* gene and is expressed mainly in cartilaginous tissue (9). Its deficiency is associated with

skeletal and chondroid disorders (10, 11). It constitutes a minor component of the extracellular matrix and plays a role in its tensile strength. COL11A1 expression is associated with poor prognosis in non-small cell lung carcinoma and linked to its resistance to chemotherapy (12). Moreover, COL11A1 promotes malignant cell proliferation and is associated with worse prognosis in many other types of cancer including ovarian cancer, squamous cell carcinoma of the head and neck (13, 14), pancreatic (15), gastric (16), colorectal (17) carcinomas, and some sarcomas (18). It is also associated with metastasis in breast cancer (19) and was previously described as a diagnostic marker to differentiate between invasive and non-invasive breast cancer (20). COL11A1 predicts invasive recurrence after primary diagnosis of breast intraductal papillomas (21), and is differentially expressed between invasive breast carcinoma and DCIS (22). However, the role of COL11A1 in DCIS and its prognostic significance has yet to be investigated. In this study, we aimed to assess the pattern of COL11A1 expression and its association with DCIS behavior in large well-annotated DCIS cohorts composed of pure DCIS and DCIS associated with invasive disease.

MATERIAL AND METHODS

Study Cohort

A large well characterized annotated cohort of DCIS including pure DCIS (n=776) and DCIS mixed with invasive breast carcinoma (DCIS-Mixed) (n=239) diagnosed between 1990 to 2012 at Nottingham City Hospital, Nottingham, United Kingdom was used in this study, as previously described (23). To avoid selection bias, the DCIS-mixed cohort was selected with clinicopathological features comparable to the pure cohort regarding the range of age at diagnosis, DCIS nuclear grade, and presence of comedo type necrosis. All demographic, clinicopathological, management data, molecular subtype, tumor infiltrating lymphocytes density and hypoxia related markers in addition to local recurrence free interval data were available and prospectively maintained (24-26). Briefly, the molecular classes were defined based on the immunohistochemistry using estrogen and progesterone receptor, the human epidermal growth factor receptor 2 and Ki-67 proliferation index. Estrogen and progesterone receptor were considered positive if >1% of tumor cells showed nuclear staining (27) while the human epidermal growth

factor receptor 2 positivity was defined when more than 10% of tumor cells showed strong complete membranous staining (+3 score). Chromogenic in situ hybridisation technique was used to determine the gene amplification status in borderline cases (28). Proliferation index was defined as low when <14% of cells showed nuclear staining (29). In addition, hypoxia inducible factor 1 alpha was evaluated using immunohistochemistry and was considered positive when >1% of tumor cells showed nuclear staining as previously described (30). Tumors with an average number of 20 lymphocytes/duct or more was considered as dense tumor infiltrating lymphocytes in DCIS as previously described (25). 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 carcinoma). Cases undergoing completion re-excision due to margins' involvement or presence of residual tumor tissue within the first 6 months were not considered as disease recurrence. Moreover, all cases with contralateral tumors or ipsilateral new event but developed in another quadrant, with morphological features different from the primary tumor or with nuclear downgrade were not considered as a recurrence and censored at time of the new event occurrence. In the pure DCIS cohort and within a median follow up period of 103 months (range 6-240), 83 (11%) cases developed recurrence [30 cases (36%) recurred as DCIS and 53 (64%) recurred as invasive disease]. The majority of the recurrences (n=66) developed in patients treated with breast conserving surgery alone.

Immunohistochemistry

Tissue microarrays were prepared from both cohorts as previously described (23). Briefly, 1 mm punch sets were used to construct the tissue microarray using an automated Grand Master 2.4-UG-EN tissue microarray machine (3DHISTECH, Budapest, Hungary). For better representation of cases with heterogeneous DCIS histological patterns and/or nuclear grade, different areas representative of such heterogeneous patterns were sampled. Prior to immunohistochemical staining, full-face tissue sections from 10 randomly selected cases were stained and assessed to evaluate the pattern of COL11A1 expression in malignant breast tissue and adjacent stroma and normal terminal duct-lobular units.

Primary antibody specificity for anti-human COL11A1 mouse monoclonal antibody, clone 1E8.33 [Ref#DMTX-0014, LOT#070912, Oncomatryx Biopharma S.L., Spain] was validated using Western blot on whole cell lysates of MCF7 and SKBR3 human breast cancer cell lines (obtained from the American Type Culture Collection; Rockville, MD, USA). COL11A1 antibody was used at a dilution of 1:1000, which showed a specific band at the predicted size of ~180 kDa.

Expression of COL11A1 protein was assessed by immunohistochemistry using the Novocastra Novolink™ Polymer Detection Systems kit (Code: RE7280-K, Leica, Biosystems, UK). Tissue microarrays and full-face sections (4 µm) were stained with mouse monoclonal COL11A1 (dilution 1:150), incubated for 24 hours after antigen retrieval using EDTA pH 9.0 in a water bath at 95°C. Hyaline cartilage tissue section in bronchi was used as positive control while the negative control was applied by omitting the primary antibody and both were simultaneously stained with the staining run.

Scoring of COL11A1 expression

Cytoplasmic expression of COL11A1 in tumor epithelial cells and the surrounding stromal cells was assessed. Semi-quantitative Histo-score was applied for cytoplasmic expression of COL11A1 in epithelial cells (staining intensity was multiplied by the percentage of representative cells in the tissue for each intensity, producing a range of values between 0 and 300) (31). The percentage of stromal fibroblasts showing cytoplasmic staining was estimated, as the intensity of staining within the scanty cytoplasm of the slender fibroblasts was difficult to assess consistently. The tumor epithelial cells and surrounding fibroblasts were assessed and scored separately. Cores containing <15% of tumor epithelial cells and associated stroma were excluded from the scoring. All scored cores showed representative areas of specialized stroma (within two high power fields) (32) surrounding the malignant duct(s). In addition, a few cores with malignant epithelial cells without surrounding stroma were excluded. The cases excluded were unbiased. Cases with multiple cores were scored and the average final score was used for the analysis. Moreover, in the DCIS-mixed cohort, each component was evaluated and scored separately for the tumor epithelial cells and surrounding stroma. All cases were scored by two pathologists (MST and IMM) using a multiheaded microscope. For dichotomization of protein expression, outcome-dependent cut-off points for either

malignant epithelial cells or stromal expression of COL11A1 were defined according to X-tile bioinformatics software (Yale University, version 3.6.1) (33) based on local recurrence free interval in the pure DCIS cohort. High COL11A1 expression within tumor epithelial cells was considered when Histo-score was >125, while expression in >15% of the surrounding fibroblasts was considered as high expression.

Analysis of *COL11A1* mRNA expression in breast cancer:

To evaluate the prognostic role of COL11A1 in breast cancer and due to the scarcity of data on the genomic and proteomic profiles of DCIS, *COL11A1* normalized mRNA expression was assessed as a potential predicative marker in the Molecular Taxonomy of Breast Cancer International Consortium cohort dataset (34), which comprises a large well characterised (n=1980) cohort of invasive breast cancer with representative molecular profile. Furthermore, to validate the prognostic role of *COL11A1* in publicly-available transcriptomic breast cancer data, analysis using the Breast Cancer Gene-Expression Miner v4.1 (bc-GenExMiner v4.1) database was carried out.

Statistical analysis

Statistical analyses were performed using SPSS v21 (Chicago, IL, USA) for Windows. Student's t-test and analysis of variance were used to correlate between *COL11A1* mRNA level and other clinicopathological parameters in METABRIC data. Association with *COL11A1* mRNA expression and breast cancer specific survival was performed after dichotomization of expression into high and low groups based on the median. Spearman's Rho test was used to correlate between COL11A1 expression within the tumor and stromal cells; as well as their correlation with tumor infiltrating lymphocytes, Ki67, and the hypoxia inducible factor 1 alpha. Association between COL11A1 expression and clinicopathological parameters in pure DCIS was performed using Chi-square, Mann Whitney U and Kruskal-Wallis tests. Wilcoxon signed rank test was used to compare the expression of COL11A1 between DCIS component and invasive component within the DCIS-mixed cases. 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 of COL11A1 expression for all recurrences (DCIS or invasive breast cancer) and invasive recurrences. For all tests, a two-tailed p -value of less than 0.05 was considered as statistically significant.

RESULTS

The pattern of COL11A1 expression

The evaluation of full-face tissue sections demonstrated even distribution of COL11A1 expression in the tumor epithelial cells and the surrounding specialized stroma throughout the whole section, indicating representability of tissue microarrays to assess its expression. Adjacent normal breast terminal ducto-lobular units showed negative or faint cytoplasmic staining. When present, COL11A1 was expressed in the cytoplasm of the epithelial tumor cells and/or surrounding fibroblasts (Figure 1). Higher stromal COL11A1 expression was observed around DCIS with periductal stromal reaction.

After unbiased exclusion of uninformative cores (lost, folded or those containing <15% tumor and surrounding stroma), the final number of cases suitable for scoring was 488 pure DCIS and 184 DCIS-mixed. COL11A1 expression showed a unimodal distribution. The median Histo-score of protein expression within the tumor was 100 in pure DCIS (range 0-160), 110 in the DCIS component of mixed cases (range 0-170), and 140 in invasive component of the latter (range 0-200). For stromal expression, the median percentage of positive stromal cells was 0% in pure DCIS (range 0-50%), 20% in the DCIS component of mixed cases (range 0-80%) and 40% in the invasive component of the latter (range 0-100%). Within the pure DCIS cohort, high COL11A1 expression was observed in 25% and 13% in tumor epithelial cells and surrounding fibroblasts; respectively. There was a positive linear correlation between expression of COL11A1 within the epithelial and surrounding stromal cells (Spearman's correlation co-efficient $r=0.3$, $p<0.0001$).

The proportion of cases with positive COL11A1 was significantly greater in DCIS-mixed than pure DCIS, both within the tumor cells (25% of pure DCIS cases vs. 40% of DCIS mixed with invasive breast cancer, $\chi^2=14.9$, $p<0.0001$) and stromal cells (13% for pure DCIS vs. 61% of DCIS mixed with invasion, $\chi^2=26.3$, $p<0.0001$). Similar results were observed when the data was analyzed using a continuous scale ($p<0.0001$ for both tumor epithelial cells and stromal cells). Moreover, there was a statistically significant

difference between COL11A1 expression within the tumor cells and surrounding stromal fibroblasts of the DCIS component and invasive component of DCIS-mixed cohort ($p < 0.0001$ for both, using categorical and continuous data) (Figure 2).

Significance of COL11A1 expression in pure DCIS

High expression of COL11A1 within the malignant epithelial cells and/or surrounding stromal fibroblasts in the pure DCIS was associated with hormonal receptor negativity, HER2 enriched and triple negative molecular subtypes. No association was observed between COL11A1 expression and the conventional clinicopathological parameters used in DCIS risk assessment such as age at diagnosis, tumor size, DCIS nuclear grade, and comedo necrosis (Table 1). Analysis of continuous data of COL11A1 expression showed similar results (Supplementary Table 1). Moreover, there was a positive linear correlation between COL11A1 expression within tumor cells and surrounding fibroblasts with tumor proliferation fraction as assessed by Ki67 ($r=0.3$; $p < 0.0001$, $r=0.14$; $p=0.01$, respectively), hypoxia inducible factor 1alpha ($r=0.15$; $p=0.03$, $r=0.13$; $p=0.01$, respectively) and tumor infiltrating lymphocytes density ($r=0.15$; $p=0.009$, $r=0.2$; $p=0.001$, respectively).

To assess the prognostic value of COL11A1 in invasive breast cancer, the Molecular Taxonomy of Breast Cancer International Consortium cohort (34) was used to correlate the levels of *COL11A1* mRNA expression with the clinicopathological variables and outcome. Higher *COL11A1* mRNA level was associated with younger age ($p < 0.0001$), lymph node metastasis ($p=0.001$), HER2 positivity ($p < 0.0001$), HER2 enriched, triple negative and luminal B subtypes in addition to shorter breast cancer specific survival ($HR=1.2$, 95% CI=1.02-1.4, $p=0.01$) (Supplementary Tables 2, and Supplementary Figure 1). Analysis using the Breast Cancer Gene-Expression Miner v4.1 (bc-GenExMiner v4.1) database showed that high *COL11A1* mRNA was associated with significantly higher metastatic relapse and shorter overall survival ($HR=1.1$, 95% CI=1.02-1.3, $p=0.03$; Supplementary Figure 2).

Outcome analysis in pure DCIS cohort

Higher COL11A1 expression within the epithelial tumor cells was associated with shorter local recurrence free interval for all recurrences (DCIS and invasive breast cancer) ($HR=2.5$; $95\%CI=1.4-4.2$, $p=0.001$) and showed a trend for invasive recurrences ($HR=1.8$; $95\%CI=0.9-3.7$, $p=0.08$) (Figure 3), while higher stromal expression was associated with shorter local recurrence free interval in the whole cohort for all recurrences and invasive recurrences ($HR=15.9$; $95\%CI=9.1-27.5$, $p<0.0001$ and ($HR=12.1$; $95\%CI=6.1-24.0$, $p<0.0001$, respectively) (Figure 4). Comparable results were obtained when the analysis was carried out with respect to different treatment groups (Figure 4). In the pure DCIS cohort, patients treated with breast conserving surgery and followed by adjuvant radiotherapy showed better outcome than those treated with breast conserving surgery only either for all recurrences ($HR=0.5$, $95\%CI=0.3-0.9$, $p=0.014$) or invasive recurrence ($HR=0.3$, $95\%CI=0.2-0.8$, $p=0.008$) (Supplementary Figure 3). Interestingly, when the cohort was stratified according to COL11A1 expression, radiotherapy did not improve the outcome in patients with high stromal COL11A1 either in all recurrences ($HR=0.8$, $95\% CI=0.5-1.6$, $p=0.587$) or for invasive recurrences ($HR=0.6$; $95\% CI=0.2-1.8$, $p=0.225$) compared to patients with low COL11A1 expression ($HR=0.2$; $95\% CI=0.2-0.9$, $p=0.04$ for all recurrences and $HR=0.4$; $95\% CI=0.1-0.7$, $p=0.03$ for invasive recurrences) (Supplementary Figure 4). Supporting this, evaluation of the interaction between high stromal COL11A1*radiotherapy using a Cox regression model showed significant association with shorter local recurrence free interval ($HR=4.2$, $95\%CI=1.5-11.7$, $p=0.006$). Furthermore, high COL11A1 expression within the stromal fibroblasts was associated with poorer outcome in all DCIS groups irrespective of the DCIS nuclear grade (low, intermediate and high) for all recurrences and for invasive recurrences (all $p<0.0001$). Supplementary Figure 5 shows forest plots for the univariate analysis of COL11A1 and other clinicopathological parameters and their association with outcome.

Multivariate Cox proportional hazards analysis for the conventional clinicopathological parameters used to assess DCIS risk showed that patient's age at diagnosis, DCIS size, nuclear grade, presence of comedo type necrosis and surgical resection margins were independent prognostic factors for disease recurrence. However, when expression of COL11A1 in stromal fibroblasts was included in the model, it was solely the independent

poor prognostic factor for tumor recurrence in patients treated with breast conserving surgery regardless of known other determinants of high risk DCIS; either for all recurrences ($HR=13.2$, $95\%CI=6.9-25.4$, $p<0.0001$) or when the analysis was confined to invasive recurrences ($HR=11.2$, $95\%CI=4.9-25.8$, $p<0.0001$), (Table 2).

Interestingly, when COL11A1 expression in the surrounding fibroblasts was incorporated with the other determinants of DCIS risk described by Van Nuys Prognostic Index (35), it provided better stratification for local recurrence risk, whereby high expression of stromal COL11A1 was associated with poorer outcome in all risk groups when compared to similar groups with low COL11A1 expression ($HR=1.9$, $95\% CI=1.2-3.1$; $p=0.004$) (Supplementary Figure 6).

DISCUSSION

Despite the advances in diagnostic and therapeutic modalities and the breakthrough in molecular genetics profiling, the underlying mechanisms promoting DCIS progression to invasive disease remain unclear and there is a desperate demand for better risk stratification tools. None of the currently available clinical indices or molecular signatures provide a reliable and clinically valid tool to predict DCIS risk of progression and/or recurrence to improve personalized management. Relying on the conventional methods for DCIS management by surgery with or without adjuvant radiotherapy for all patients based on such inadequate risk assessment leads to over- or under-treatment of a substantial proportion of patients. Furthermore, the biological and clinical heterogeneity of DCIS makes risk stratification quite challenging. Studying the role of DCIS microenvironment and the interaction between its various components and understanding how this affects disease behavior could resolve the DCIS dilemma and provide an adequate risk stratification model (36-39). As crosstalk between tumor epithelial cells and extracellular matrix including collagen remodelling, deposition and degradation is an essential key step in carcinogenesis and the progression of *in situ* tumors to invasive disease, studying potential markers that drive this process and their prognostic value is a credible approach to refine DCIS risk.

COL11A1 is a minor fibrillary collagen present mainly in cartilage, but it is produced at varying levels by other normal tissues as well as and under some pathological conditions. Interestingly, several studies reported that overexpression of COL11A1 by tumor cells and/or the surrounding cancer associated fibroblasts is linked with tumor progression and poorer prognosis in a variety of cancers including invasive breast cancer (12-14, 16, 17, 40). Moreover, in invasive breast cancer, some authors reported that COL11A1 is exclusively produced by invasive tumors (20), while others showed that it is expressed in normal, preinvasive and invasive tumors at different levels (40). *COL11A1* is differentially expressed between invasive breast cancer and DCIS (22), however there is no previous study investigating its role in DCIS progression and its potential prognostic significance. Furthermore, using the Molecular Taxonomy of Breast Cancer International Consortium cohort for robust molecular data in a large number of invasive breast cancer, we have shown an association between aggressive behavior of invasive

breast cancer and higher levels of *COL11A1* mRNA. Taken together, these observations support our hypothesis that *COL11A1* is a promising potential protein that could provide additional knowledge to DCIS behavior and might help in the stratification of disease risk.

Here we explored the expression of *COL11A1* in a large well characterized cohort of DCIS and scored the protein expression in tumor cells and surrounding stromal fibroblasts. Interestingly, high *COL11A1* expression was associated with some features of high-risk DCIS, supporting its role in DCIS progression. Additionally, our data showed that *COL11A1* expression was higher in DCIS co-existing with invasive breast cancer than pure DCIS, and much higher in the invasive component both within the tumor cells and in the surrounding stroma.

The prognostic value of *COL11A1* was shown with a shorter local recurrence free interval in patients with higher levels of *COL11A1* expression, particularly stromal expression, independently from other clinicopathological factors. These findings were consistent for all recurrent events including DCIS and invasive breast cancer, and also when the analysis was confined to invasive recurrences only, which provides more evidence that *COL11A1* could play a key role in DCIS progression to invasive disease. Although it was thought that *COL11A1* is exclusively produced by cancer associated fibroblasts, other studies showed that *COL11A1* was expressed in tumor cells as well as the surrounding stroma (22, 40). Our study showed that expression of *COL11A1* in tumor epithelial and stromal cells is associated with disease recurrence; a finding that might reflect the epithelial-stromal interaction and its role in progression of DCIS. This interaction is supported by the obvious increase of *COL11A1* expression in stromal and epithelial cells surrounding the invasive component compared to those surrounding the DCIS component in mixed cases or those surrounding pure DCIS. In our cohort, we have observed that higher expression *COL11A1* is associated with periductal stromal reaction around DCIS, a finding that was reported to be associated with invasive recurrence in DCIS (41). However, further functional studies are highly recommended to understand the underlying mechanisms and functions of *COL11A1* expression in carcinogenesis and tumor progression either from the tumor cells or the surrounding stroma.

Multiple authors have noted a group of low grade DCIS with indolent appearance and low proliferation index that yet carries progression potential to invasive breast cancer (42-44). An explanation of disease progression based exclusively on intrinsic epithelial tumor cell-related factors is inadequate and identifying other candidate potential markers that could explain progression of different phenotypes and hence better risk stratification is highly warranted. Our findings showed that there was no association between high COL11A1 expression and other conventional clinicopathological factors that are described as a surrogate for high risk DCIS such as younger age at diagnosis, high grade DCIS, comedo type necrosis and mode of presentation. In addition, high expression of stromal COL11A1 showed an association with recurrence regardless the grade of the DCIS or method of therapy. These findings show that COL11A1 is a very promising protein for better DCIS risk stratification and understanding of the disease biology. Moreover, incorporation of COL11A1 with the other clinicopathological factors described in the Van Nuys Prognostic Index provided a better stratification of different risk groups. These findings indicate that COL11A1 is a marker that could be used to better define high-risk DCIS and identify of patients who need more radical treatment e.g. breast conserving surgery with wider excision margins or mastectomy.

Although the effect of radiotherapy in reduction of DCIS recurrence is undeniable (45-48), there is still a considerable proportion of patients treated with breast conserving surgery and followed by adjuvant radiotherapy who develop disease recurrence. There are no data available for DCIS patients at higher susceptibility to show poor outcome after adjuvant radiotherapy management. Our findings showed that high expression of stromal COL11A1 is associated with higher rate of recurrence even with radiotherapy compared to DCIS with lower levels of expression, which indicates that COL11A1 could provide a signature for radiotherapy resistance in DCIS. However, the number of cases with recurrence after radiotherapy, especially invasive recurrence, is low, and this finding needs to be assessed in an independent cohort. Studies showed that dense stroma is associated with poor response to radiotherapy in cancer treatment (49, 50) which might provide a similar explanation in DCIS. These findings warrant more validation and mechanistic studies to decipher the underlying mechanisms and biology.

Our data support a key role for COL11A1 in progression to invasive disease and yet high levels of COL11A1 were highly significantly related to invasive recurrence. Thus far, understanding of the biological processes which involve COL11A1 in cancer progression is lacking, although we hypothesize that high stromal COL11A1 may in some way enable individual tumor cell invasion and/or remnant cell survival beyond the surgical margin, leading to invasive recurrence. Collagen provides a scaffold for the tumor cells and helps in tumor growth, angiogenesis and invasion (8). These findings are supported by the association between COL11A1 and proliferation and hypoxia related proteins in our cohort. The link between hypoxia and increased collagen deposition and remodelling through prolyl-hydroxylases and lysyl oxidase enzymes is well studied even in DCIS (51-53). One of the potential mechanisms that COL11A1 could use in promoting cellular proliferation is its association with the Smad signalling pathway via binding to integrin (12). In addition, it was thought that collagen acts as a barrier against tumor invasion, but now it is proven that it has multifaceted roles and could promote tumor progression. Increased collagen deposition promotes tumor progression through destabilization of cell polarity and cell-cell adhesion, and enhancement of growth factor signaling. Tissue tension regulates the epithelial-mesenchymal transition through crosslink between tumor cells and stromal cells (54, 55).

Although the role of collagen in inflammatory processes and tissue healing is obvious, the specific role of COL11A1 in such conditions is unclear. Our data showed a correlation between higher COL11A1 expression mainly in stromal cells and dense inflammatory cell infiltrates. The underlying biology is warranted to be investigated, as we previously reported that dense tumor infiltrating lymphocytes have poor prognostic significance in DCIS, a phenomenon different from invasive disease and for which the underlying mechanisms are unclear (25). COL11A1 may interact with the inflammatory related markers; for instance, cytokines and or interleukins, and affect DCIS behavior.

Conclusion

Extracellular matrix remodelling is an essential feature associated with DCIS that could lead to either promoting or circumventing its progression to invasive disease. COL11A1 might have a potential role in DCIS aggressiveness through its collagen remodeling and

regulatory mechanism in cellular proliferation. Additional functional studies to decipher the role of COL11A1 and its mechanism of action in DCIS behavior are indicated. Applying COL11A1 in clinical practice, especially stromal expression, would provide a highly promising prognostic indicator for DCIS invasive recurrence and may be a potential predictive marker for radiotherapy response. As a consequence, the group of patients with high stromal COL11A1 (13% of cases) could require more extensive surgery, regardless of other factors.

Limitations of the study

This study has been carried out on tissue microarray sections, which might underestimate the role of tumor heterogeneity. However, all cases in our cohort were histologically reviewed before tissue microarray construction and we used multiple cores for cases with heterogeneous grades or morphological patterns. Moreover, our cohort did not include any patients treated with endocrine therapy, and the number of recurrences after radiotherapy was small. Further validation studies on larger cohorts especially with radiotherapy treated patients is highly warranted.

Ethics:

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.

Acknowledgment:

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

References

1. Rakovitch E, Nofech-Mozes S, Hanna W, Baehner FL, Saskin R, Butler SM, *et al.* A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast Cancer Res Treat.* 2015;152:389-98.
2. 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.
3. Asjoe FT, Altintas S, Huizing MT, Colpaert C, Marck EV, Vermorken JB, *et al.* The value of the Van Nuys Prognostic Index in ductal carcinoma in situ of the breast: a retrospective analysis. *Breast J.* 2007;13:359-67.
4. 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.
5. 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.
6. Mazouni C, Delalogue 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.
7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646-74.
8. Prockop DJ, Kivirikko KI, Tuderman L, Guzman NA. The biosynthesis of collagen and its disorders (first of two parts). *N Engl J Med.* 1979;301:13-23.
9. Bernard M, Yoshioka H, Rodriguez E, Van der Rest M, Kimura T, Ninomiya Y, *et al.* Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA demonstrates that type XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilagenous tissue. *J Biol Chem.* 1988;263:17159-66.
10. Tompson SW, Bacino CA, Safina NP, Bober MB, Proud VK, Funari T, *et al.* Fibrochondrogenesis results from mutations in the COL11A1 type XI collagen gene. *Am J Hum Genet.* 2010;87:708-12.
11. Khalifa O, Imtiaz F, Allam R, Al-Hassnan Z, Al-Hemidan A, Al-Mane K, *et al.* A recessive form of Marshall syndrome is caused by a mutation in the COL11A1 gene. *J Med Genet.* 2012;49:246-8.
12. Shen L, Yang M, Lin Q, Zhang Z, Zhu B, Miao C. COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. *Oncol Rep.* 2016;36:877-85.
13. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* 2014;33:3432-40.

14. Sok JC, Lee JA, Dasari S, Joyce S, Contrucci SC, Egloff AM, *et al.* Collagen type XI alpha1 facilitates head and neck squamous cell cancer growth and invasion. *Br J Cancer.* 2013;109:3049-56.
15. Erkan M, Weis N, Pan Z, Schwager C, Samkharadze T, Jiang X, *et al.* Organ-, inflammation- and cancer specific transcriptional fingerprints of pancreatic and hepatic stellate cells. *Mol Cancer.* 2010;9:88.
16. Zhao Y, Zhou T, Li A, Yao H, He F, Wang L, *et al.* A potential role of collagens expression in distinguishing between premalignant and malignant lesions in stomach. *Anat Rec (Hoboken)* 2009;292:692-700.
17. Fischer H, Stenling R, Rubio C, Lindblom A. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. *Carcinogenesis* 2001;22:875-8.
18. Kleman JP, Hartmann DJ, Ramirez F, van der Rest M. The human rhabdomyosarcoma cell line A204 lays down a highly insoluble matrix composed mainly of alpha 1 type-XI and alpha 2 type-V collagen chains. *Eur J Biochem* 1992;210:329-35.
19. Ellsworth RE, Seebach J, Field LA, Heckman C, Kane J, Hooke JA, *et al.* A gene expression signature that defines breast cancer metastases. *Clin Exp Metastasis* 2009;26:205-13.
20. Freire J, Dominguez-Hormaetxe S, Pereda S, De Juan A, Vega A, Simon L, *et al.* Collagen, type XI, alpha 1: an accurate marker for differential diagnosis of breast carcinoma invasiveness in core needle biopsies. *Pathol Res Pract.* 2014;210:879-84.
21. Freire J, Garcia-Berbel L, Garcia-Berbel P, Pereda S, Azueta A, Garcia-Arranz P, *et al.* Collagen Type XI Alpha 1 Expression in Intraductal Papillomas Predicts Malignant Recurrence. *BioMed Res Int.* 2015;2015:812027.
22. Vargas AC, McCart Reed AE, Waddell N, Lane A, Reid LE, Smart CE, *et al.* Gene expression profiling of tumor epithelial and stromal compartments during breast cancer progression. *Breast Cancer Res Treat.* 2012;135:153-65.
23. Miligy IM, Gorringer 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(12):1807-15.
24. Toss MS, Miligy IM, Gorringer 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.
25. 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.
26. Toss MS, Miligy IM, Gorringer 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.

27. 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.
28. 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.
29. 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. *Ann Oncol* 2011;22:1736-47.
30. 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.
31. McCarty KS, McCarty KS. Histochemical approaches to steroid receptor analyses. *Semin Diagn Pathol* 1984;1:297-308.
32. Toss M, Miligy I, Thompson AM, Khout H, Green AR, Ellis IO, *et al.* Current trials to reduce surgical intervention in ductal carcinoma in situ of the breast: Critical review. *Breast* 2017;35:151-6.
33. 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.
34. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346-52.
35. Boland GP, Chan KC, Knox WF, Roberts SA, Bundred NJ. Value of the Van Nuys Prognostic Index in prediction of recurrence of ductal carcinoma in situ after breast-conserving surgery. *Br J Surg* 2003;90:426-32.
36. 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 alpha6beta6 integrin in DCIS identifies high-risk patients and predicts recurrence. *Clin Cancer Res* 2014;20:344-57.
37. Semenza GL. The hypoxic tumor microenvironment: A driving force for breast cancer progression. *Biochim Biophys Acta* 2016;1863:382-91.
38. 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.
39. Allinen M, Beroukhi 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.

40. Halsted KC, Bowen KB, Bond L, Luman SE, Jorcyk CL, Fyffe WE, *et al.* Collagen alpha1(XI) in normal and malignant breast tissue. *Mod Pathol* 2008;21:1246-54.
41. Visser LL, Elshof LE, Schaapveld M, van de Vijver K, Groen EJ, Almekinders MM, *et al.* Clinicopathological Risk Factors for an Invasive Breast Cancer Recurrence after Ductal Carcinoma In Situ-A Nested Case-Control Study. *Clin Cancer Res* 2018;24:3593-601.
42. Cowell CF, Weigelt B, Sakr RA, Ng CK, Hicks J, King TA, *et al.* Progression from ductal carcinoma in situ to invasive breast cancer: revisited. *Mol Oncol* 2013;7:859-69.
43. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res* 2009;11:R7.
44. Doebar SC, Krol NM, van Marion R, Brouwer RWW, van Ijcken WFJ, Martens JM, *et al.* Progression of ductal carcinoma in situ to invasive breast cancer: comparative genomic sequencing. *Virchows Arch.* 2019;474:247-51.
45. 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.
46. 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.
47. 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.
48. 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.
49. Park CC, Rembert J, Chew K, Moore D, Kerlikowske K. High Mammographic Breast Density Is Independent Predictor of Local But Not Distant Recurrence After Lumpectomy and Radiotherapy for Invasive Breast Cancer. *Int J Radiat Oncol Biol Phys.* 2009;73:75-9.
50. Barker HE, Paget JTE, Khan AA, Harrington KJ. The tumour microenvironment after radiotherapy: mechanisms of resistance and recurrence. *Nat Rev Cancer* 2015;15:409-25.
51. Toss MS, Miligy IM, Gorringer 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(12):1518-1526

52. Bos R, van Diest PJ, de Jong JS, van der Groep P, van der Valk P, van der Wall E. Hypoxia-inducible factor-1alpha is associated with angiogenesis, and expression of bFGF, PDGF-BB, and EGFR in invasive breast cancer. *Histopathology* 2005;46:31-6.
53. 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.
54. Wolf K, Wu YI, Liu Y, Geiger J, Tam E, Overall C, *et al.* Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. *Nat Cell Biol.* 2007;9:893-904.
55. Motrescu ER, Blaise S, Etique N, Messaddeq N, Chenard MP, Stoll I, *et al.* Matrix metalloproteinase-11/stromelysin-3 exhibits collagenolytic function against collagen VI under normal and malignant conditions. *Oncogene* 2008;27:6347-55.

Table 1: Correlation between COL11A1 expression with different clinicopathological parameters in the pure DCIS cohort.

Clinicopathological Parameters	COL11A1 expression in tumor epithelial cells		χ^2 (p-value)	COL11A1 expression in stromal fibroblasts		χ^2 (p-value)
	Low (N=338) N. (%)	High (N=110) N. (%)		Low (N=391) N. (%)	High (N=57) N. (%)	
Age (years) ≤50 >50	81 (24) 257 (76)	27 (25) 83 (75)	0.1 (0.902)	91 (23) 300 (77)	17 (30) 40 (70)	1.2 (0.280)
Presentation Screening Symptomatic	166 (49) 172 (51)	52 (47) 58 (53)	0.1 (0.737)	193 (49) 198 (51)	25 (44) 32 (56)	0.6 (0.438)
DCIS Size (mm) ≤20 >20	147 (44) 191 (56)	51 (47) 57 (53)	0.4 (0.497)	163 (43) 223 (57)	30 (54) 25 (46)	2.6 (0.106)
Nuclear Grade Low Moderate High	40 (12) 88 (26) 210 (62)	13 (12) 31 (28) 66 (60)	0.2 (0.903)	49 (13) 102 (26) 240 (61)	4 (7) 17 (30) 36 (63)	1.6 (0.457)
Comedo necrosis Yes No	221 (65) 117 (35)	76 (69) 34 (31)	0.5 (0.475)	258 (66) 133 (34)	39 (68) 18 (32)	0.1 (0.716)
Estrogen receptor Negative Positive	77 (25) 233 (75)	34 (33) 69 (67)	2.6 (0.105)	87 (24) 272 (76)	24 (44) 30 (56)	9.8 (0.002)
Progesterone receptor Negative Positive	119 (38) 192 (62)	53 (52) 48 (48)	6.3 (0.012)	144 (40) 217 (60)	28 (55) 23 (45)	4.1 (0.042)
HER2 status Negative Positive	231 (76) 74 (24)	75 (74) 27 (26)	0.2 (0.655)	289 (77) 85 (23)	39 (70) 17 (30)	1.5 (0.211)
Proliferation index (Ki-67) High Low	62 (23) 210 (77)	25 (24) 78 (76)	0.1 (0.762)	70 (22) 255 (78)	17 (34) 33 (66)	3.8 (0.049)
Molecular classes Luminal A Luminal B HER2 Enriched Triple negative	139 (54) 47 (18) 40 (15) 34 (13)	47 (48) 18 (18) 17 (17) 17 (17)	1.5 (0.685)	167 (54) 57 (18) 43 (14) 42 (14)	19 (38) 8 (16) 14 (28) 9 (18)	8.2 (0.041)
Tumor infiltrating lymphocytes Dense Sparse	122 (47) 137 (53)	49 (57) 37 (43)	2.5 (0.113)	140 (47) 158 (53)	31 (66) 16 (34)	5.9 (0.016)
HIF1A expression High Low	58 (24) 188 (76)	26 (29) 63 (71)	1.1 (0.293)	63 (21) 228 (78)	21 (48) 23 (52)	13.8 (<0.0001)

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

COL11A1; Collagen (XI) alpha-1 chain, DCIS; ductal carcinoma in situ, HER2; Human epidermal growth factor receptor 2, HIF1 α ; hypoxia-inducible factor 1 alpha.

Table 2: Multivariate survival analysis (Cox regression model) of variables (with and without COL11A1) predicting outcome in terms of ipsilateral local all recurrences (A) and invasive recurrences (B) in patients treated by breast conserving surgery in pure DCIS.

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
COL11A1 and other clinicopathological parameters associated with high risk DCIS				
High stromal COL11A1 expression	13.2	6.9	25.4	<0.0001
Patient Age	0.6	0.3	1.4	0.261
DCIS presentation	1.9	0.9	3.9	0.084
DCIS size	1.1	0.7	1.9	0.609
DCIS nuclear Grade	1.6	0.9	3.1	0.123
Comedo necrosis	0.5	0.2	1.2	0.120
Margin status	1.0	0.9	1.1	0.728

Significant *p* values are in **bold**
 COL11A1; Collagen (XI) alpha-1 chain, DCIS; ductal carcinoma in situ.

B) Invasive recurrence

Conventional clinicopathological parameters associated with high risk DCIS				
Parameters	Hazard ratio (<i>HR</i>)	95% <i>confidence interval (CI)</i>		<i>p</i> -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
COL11A1 and other clinicopathological parameters associated with high risk DCIS				
High stromal COL11A1 expression	11.2	4.9	25.8	<0.0001
Patient Age	0.9	0.3	3.3	0.974
DCIS presentation	2.0	0.8	5.2	0.156
DCIS size	1.5	0.7	2.9	0.258
DCIS nuclear Grade	1.9	0.8	4.4	0.119
Comedo necrosis	0.5	0.2	1.3	0.159
Margin status	1.1	0.9	1.2	0.386

Significant *p* values are in **bold**
 COL11A1; Collagen (XI) alpha-1 chain, DCIS; ductal carcinoma in situ.

Supplementary Table 1: Correlation between COL11A1 expression and different clinicopathological parameters in the pure DCIS cohort using continuous data.

Clinicopathological Parameters	Number of cases	COL11A1 expression in tumor epithelial cells		COL11A1 expression in stromal fibroblasts	
		Mean Rank	<i>p</i> -value	Mean Rank	<i>p</i> -value
Age (years) ≤50 >50	108 340	225.8 224.1	0.904	233.4 221.7	0.296
Presentation Screening Symptomatic	218 230	222.8 226.1	0.789	219.5 229.2	0.312
DCIS Size (mm) ≤20 >20	198 248	223.2 223.7	0.965	232.4 216.3	0.094
Nuclear Grade Low Moderate High	53 119 276	223.1 224.0 225.0	0.993	201.6 214.8 233.1	0.056
Comedo necrosis Yes No	297 151	215.6 229.1	0.295	226.7 220.2	0.523
Estrogen receptor Negative Positive	111 302	244.7 193.15	<0.0001	236.9 196.0	<0.0001
Progesterone Receptor Negative Positive	172 240	233.8 186.9	<0.0001	222.2 195.3	0.004
HER2 status Negative Positive	306 101	198.0 222.1	0.073	210.6 231.4	0.061
Proliferation index (Ki-67) High Low	87 288	215.6 179.6	0.006	203.6 183.6	0.055
Molecular classes Luminal A Luminal B HER2 Enriched Triple negative	186 65 57 51	166.4 179.1 215.8 190.7	0.014	167.9 177.4 212.7 190.9	0.004
Tumor infiltrating lymphocytes Dense Sparse	171 174	189.1 157.1	0.003	187.1 159.1	0.001
HIF1A expression High Low	84 251	192.2 159.9	0.008	192.3 159.8	0.001

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

COL11A1; Collagen (XI) alpha-1 chain, DCIS; ductal carcinoma in situ, HER2; Human epidermal growth factor receptor 2, HIF1 α ; hypoxia inducible factor 1 alpha.

Supplementary Table 2: Correlation between COL11A1 mRNA level and the clinicopathological parameters in the METABRIC series of invasive breast cancers (n=1980).

Clinicopathological parameters	Number of cases	Mean <i>COL11A1</i> mRNA level	<i>p</i> -value
Patient Age (years) <50 ≥50	383 1556	8.7 8.8	<0.0001
Tumor Size (mm) ≤20 >20	622 1331	8.7 8.7	0.803
Histologic Grade 1 2 3	170 770 952	8.8 8.9 8.9	0.621
Lymph node Stage 1 2 3	1035 622 316	8.7 8.9 8.8	0.001
Estrogen Receptor Status Positive Negative	1506 474	8.7 8.7	0.163
HER2 Status Negative Positive	1733 247	8.7 8.8	<0.0001
PAM50 molecular classes Luminal A Luminal B Basal-like HER2 enriched Normal like	718 488 329 240 199	8.7 8.9 8.9 8.7 8.5	<0.0001

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

METABRIC; Molecular Taxonomy of Breast Cancer International Consortium

Figures:

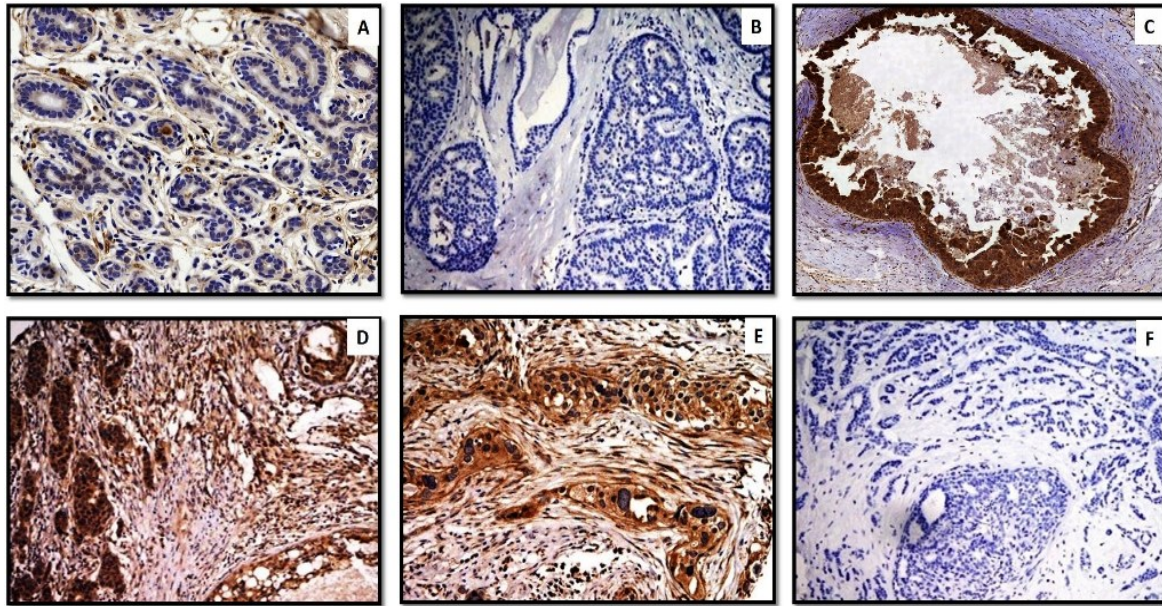


Figure 1: Representative photomicrographs of COL11A1 IHC expression, A) Normal breast ductalobular unit (x20) shows negative staining of COL11A1 in epithelial cells and occasional staining in the surrounding fibroblasts; B) Negative COL11A1 expression (x20) in a pure DCIS case; C) strong expression of COL11A1 in tumor cells and surrounding fibroblasts (x20) in a pure DCIS case; D and E) Expression of COL11A1 in a mixed case (x40) showing strong staining in invasive component within the tumor cells and/or surrounding stromal fibroblasts; and F) negative expression of COL11A1 in a mixed case (x20).

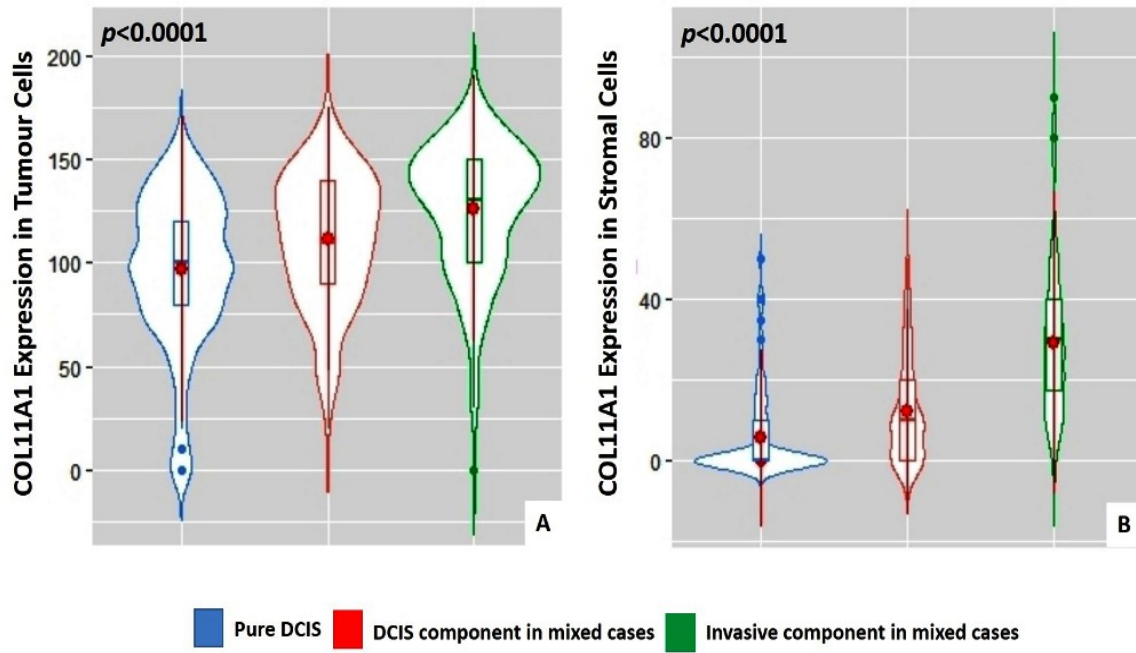


Figure 2: Violin plots showing differences of COL11A1 expression between pure DCIS and DCIS-mixed both in tumor cells (A) and surrounding stroma (B).

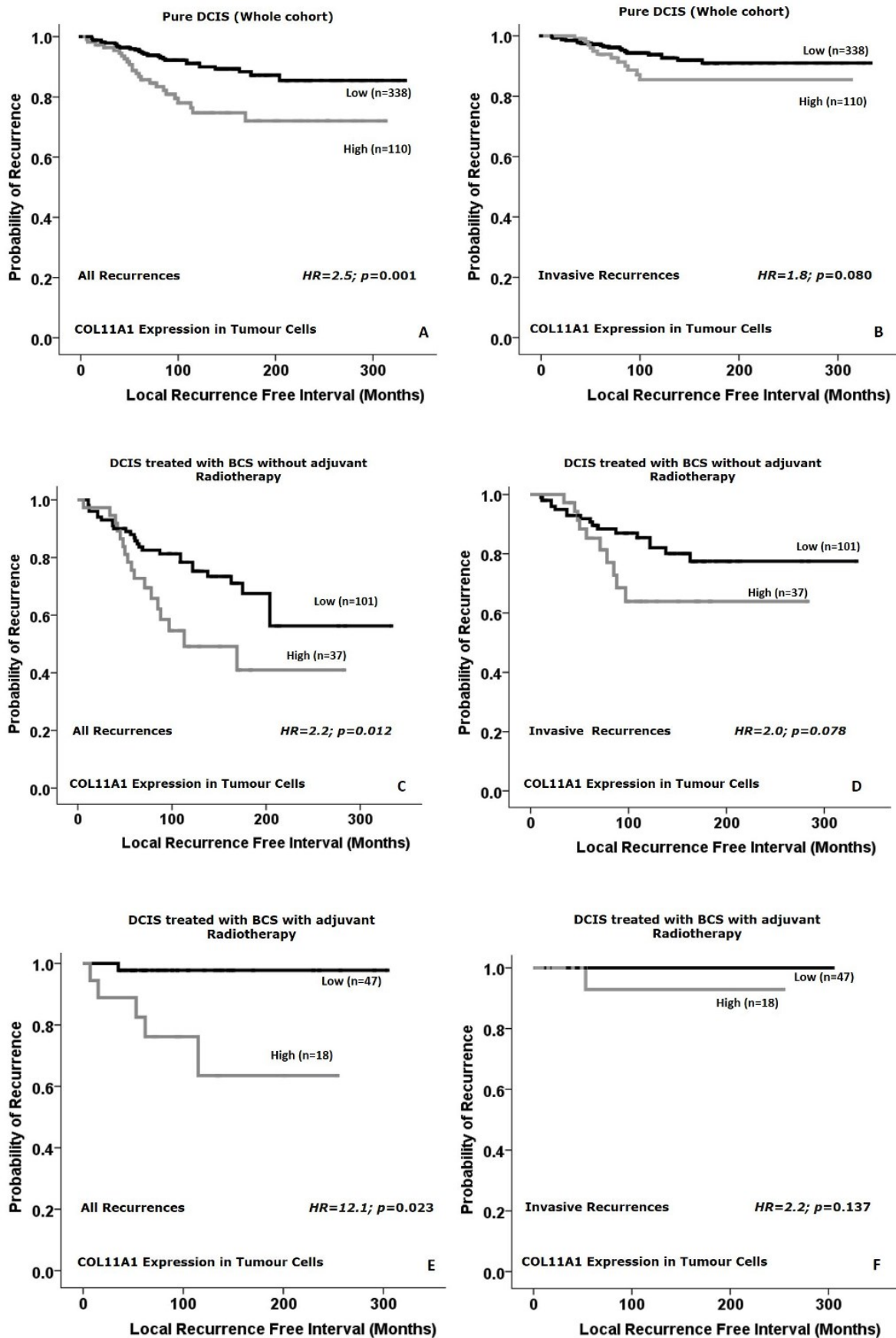


Figure 3: Kaplan Meier curves showing that high expression of COL11A1 within the tumor epithelial cells is associated with ipsilateral local recurrence free interval in the whole series either for all recurrences (A) or invasive recurrences (B), in breast conserving surgery without adjuvant

radiotherapy (C; all recurrences and D; invasive recurrences) and in patients treated with breast conserving surgery followed by adjuvant radiotherapy (E; all recurrences and F; invasive recurrences).

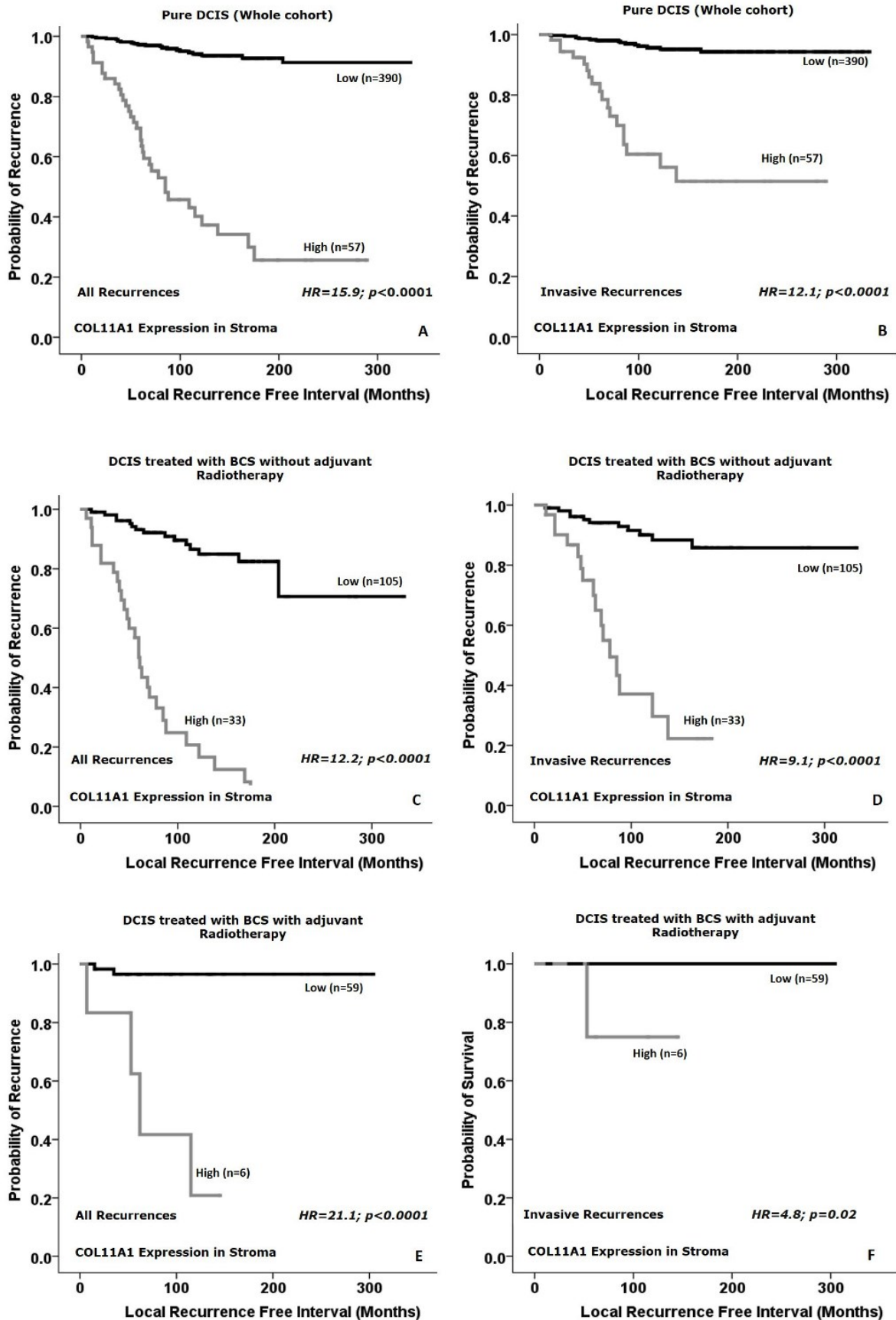
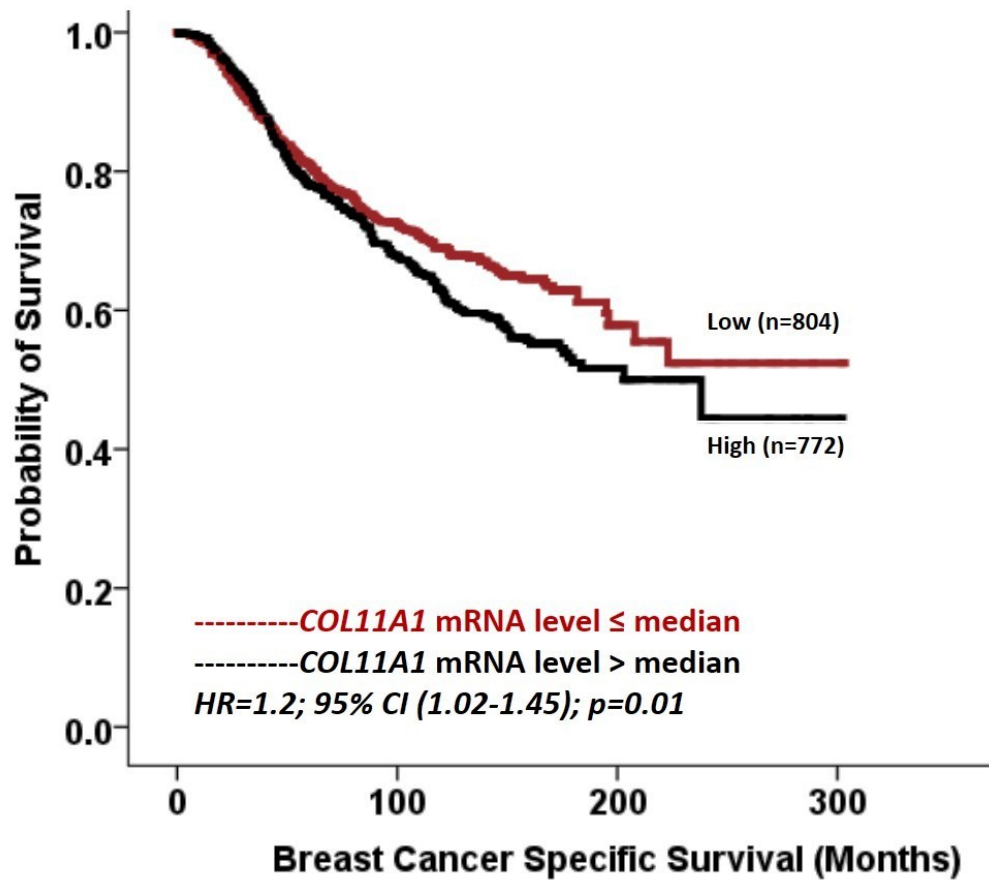


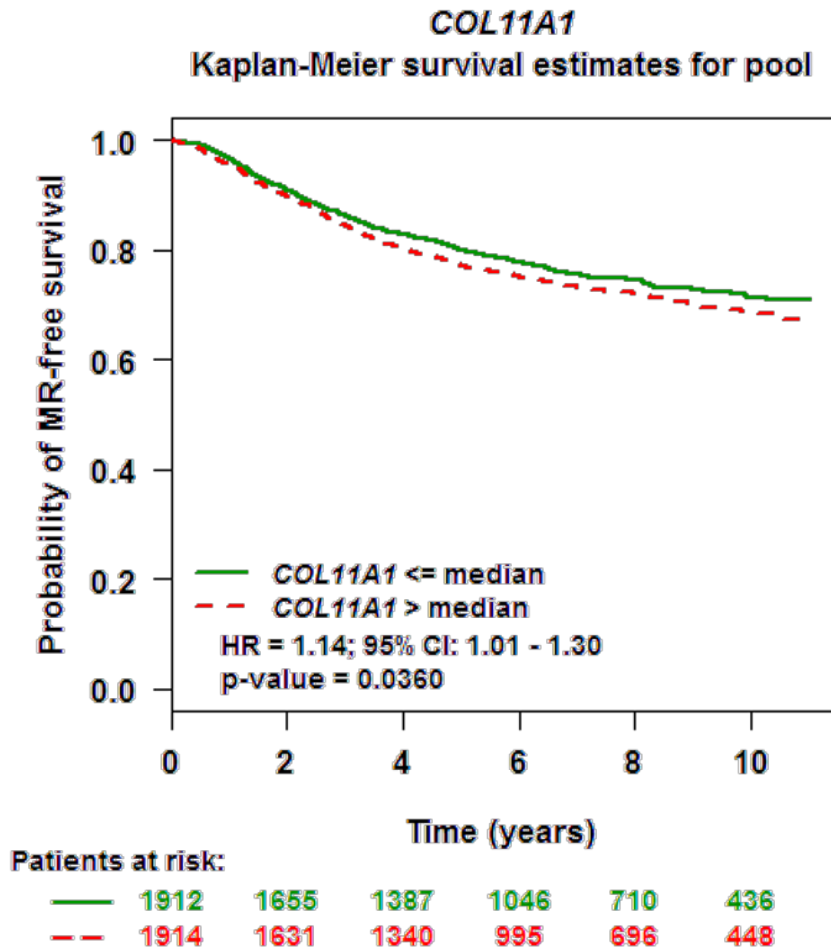
Figure 4: Kaplan Meier curves showing that high expression of COL11A1 within the surrounding stromal fibroblast is associated with ipsilateral local recurrence free interval in the whole series either for all recurrences (A) or invasive recurrences (B), in breast conserving surgery without adjuvant

radiotherapy (C; all recurrences and D; invasive recurrences) and in patients treated with breast conserving surgery followed by adjuvant radiotherapy (E; all recurrences and F; invasive recurrences).

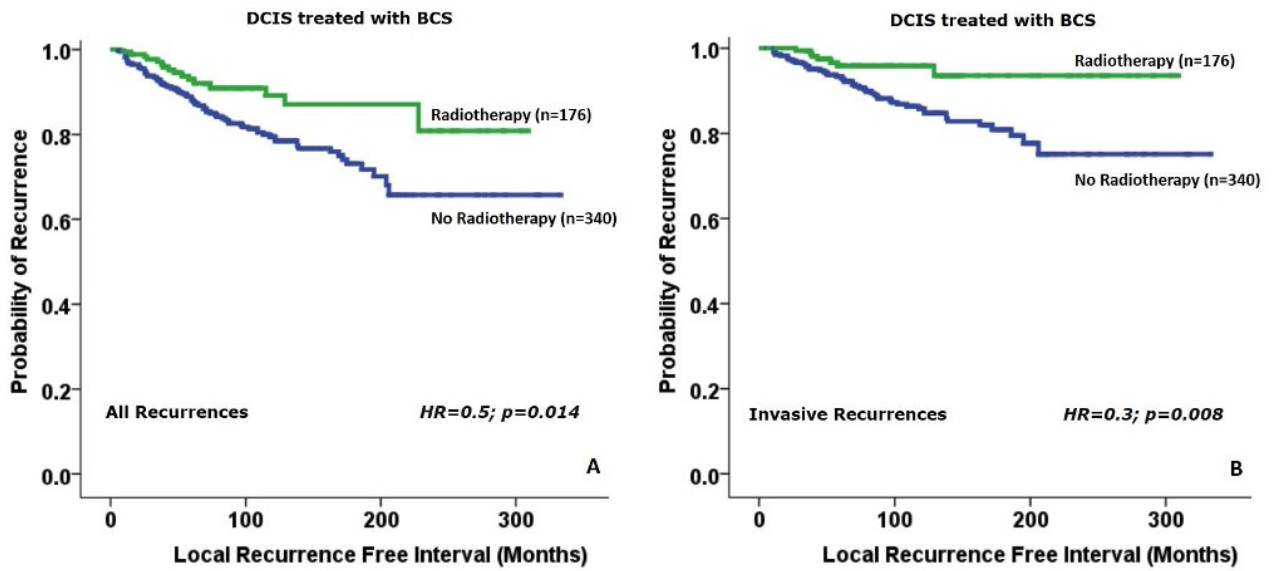
Supplementary Figures:



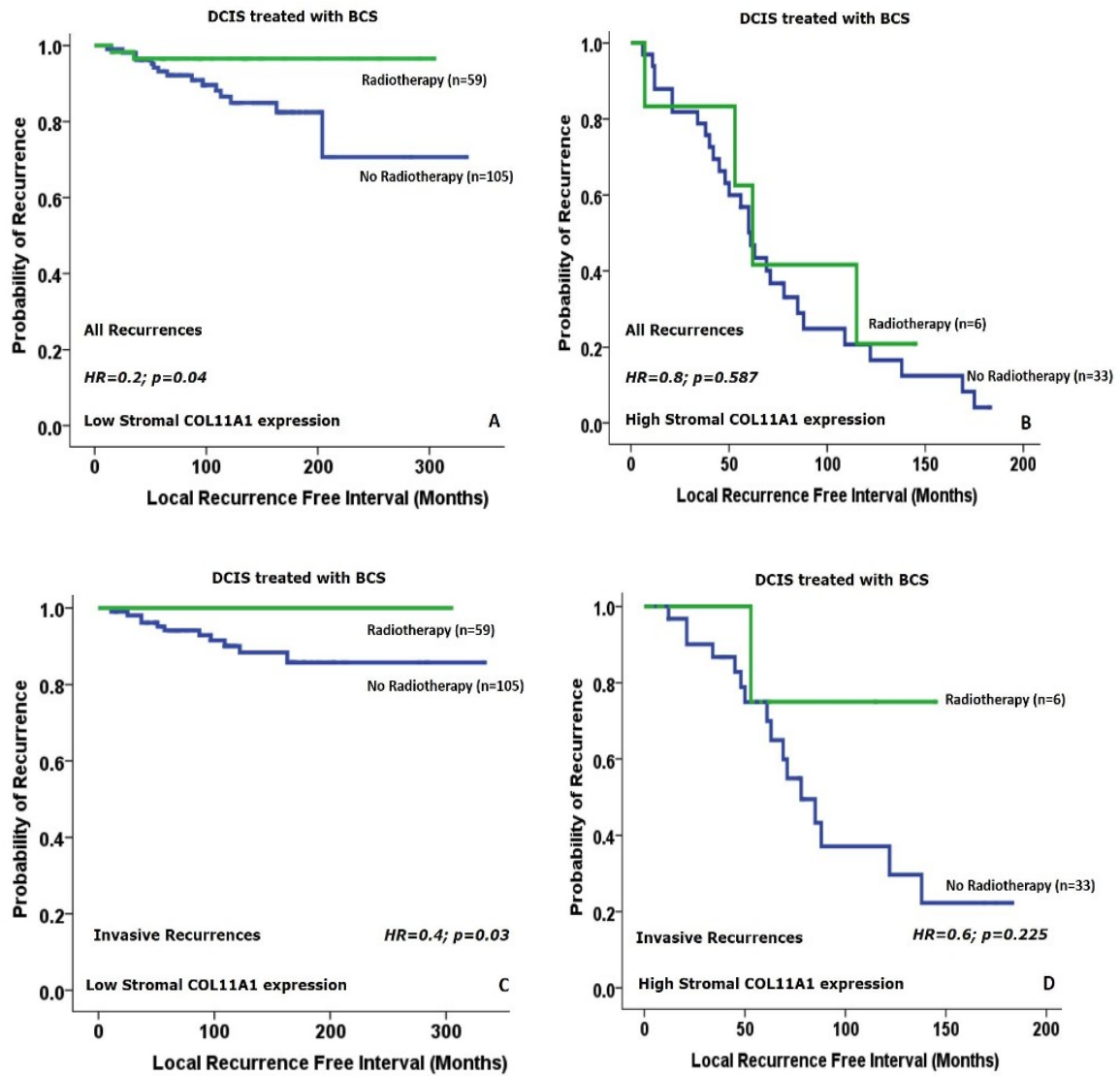
Supplementary Figure 1: Association between *COL11A1* mRNA level and outcome in terms of breast cancer specific survival in the METABRIC series. The cohort was split into high and low mRNA expression at the median mRNA expression.



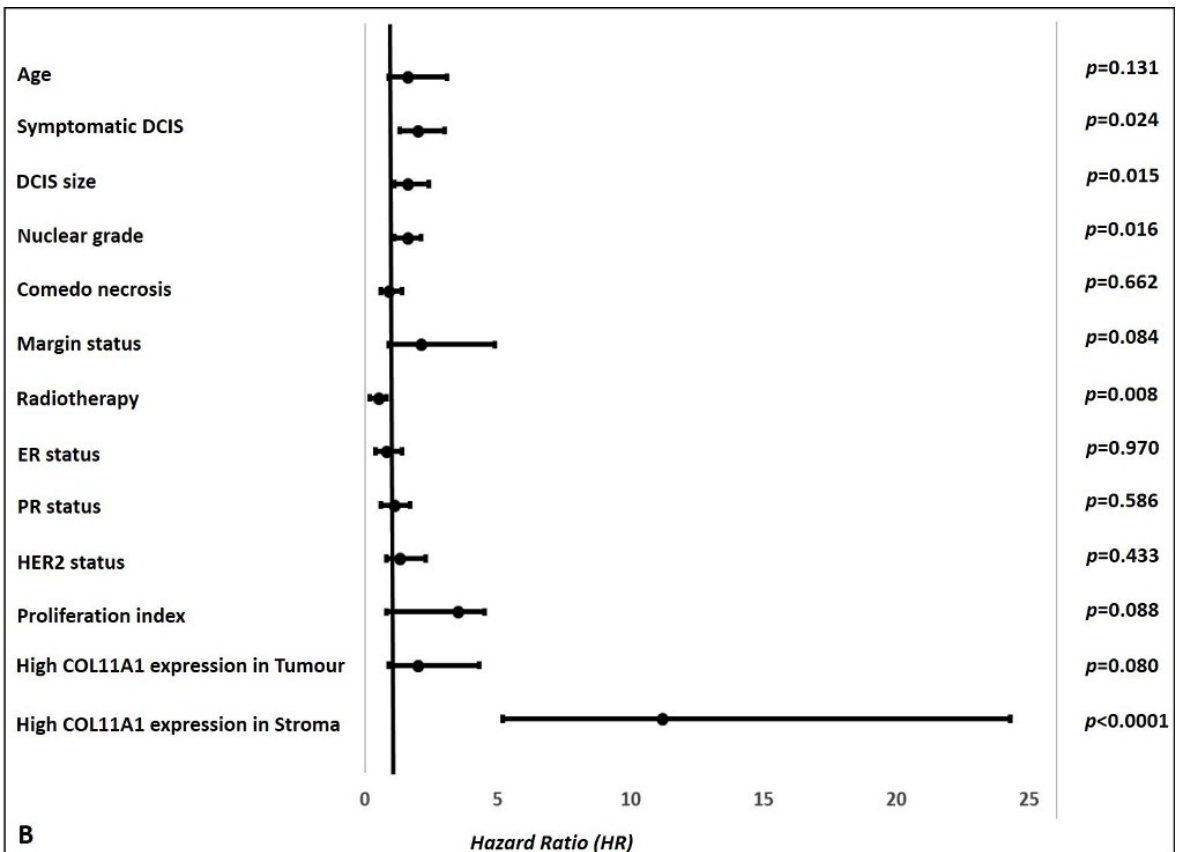
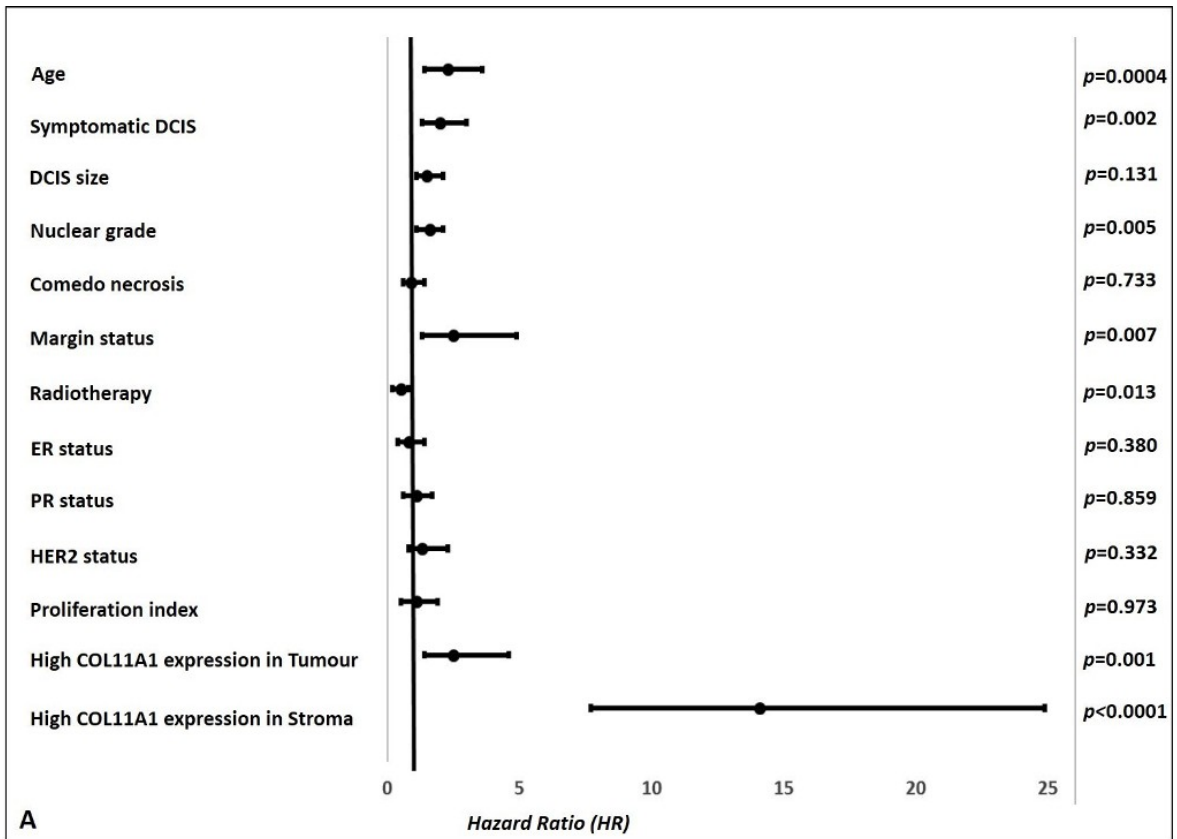
Supplementary Figure 2: Association between *COL11A1* mRNA level and outcome (Metastasis free survival) in Breast Cancer Gene-Expression Miner database.



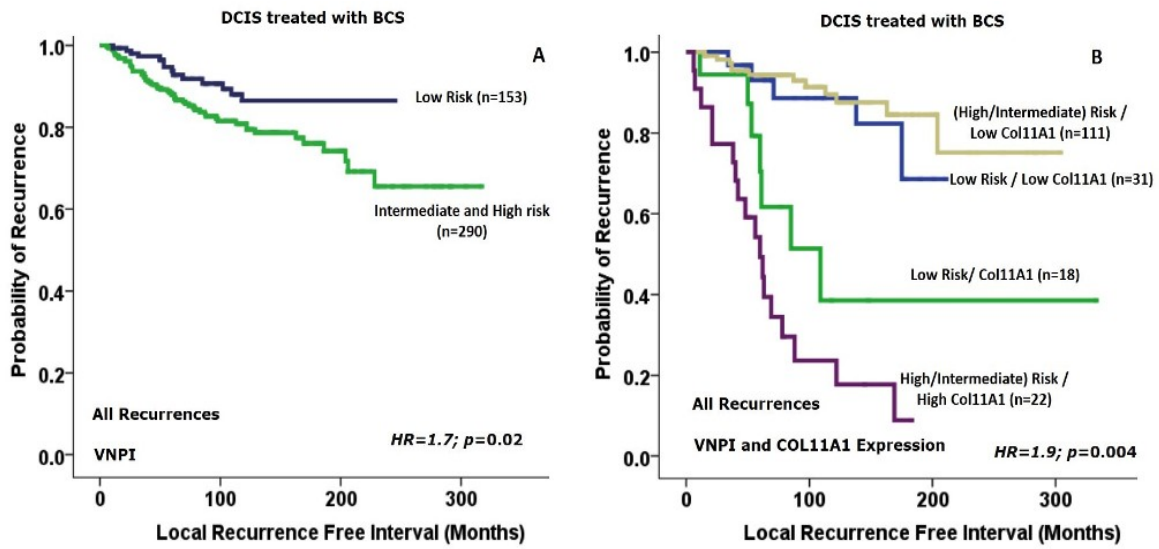
Supplementary Figure 3: Kaplan Meier curves showing the association between radiotherapy and better outcome in the pure DCIS cohort for all recurrences (A) and invasive recurrences (B).



Supplementary Figure 4: Kaplan Meier curves showing that radiotherapy is associated with better outcome in patients treated with breast conserving surgery and have low stromal COL11A1 expression for all recurrences (A), and invasive recurrences (C). While patients with high stromal expression of COL11A1 showed no difference in recurrence rate after radiotherapy for all recurrences (B), and for invasive recurrences (D).



Supplementary Figure 5: Forest plots showing the univariate analysis results of association between different clinicopathological parameters and ipsilateral tumor recurrence for patients treated with breast conserving surgery in pure DCIS cohort; A) all recurrences whether DCIS or invasive and B) for invasive recurrences only.



Supplementary Figure 6: Kaplan Meier curves showing the association between DCIS risk and local recurrence free interval in patients treated with breast conserving surgery based on Van Nuys Prognostic Index alone (A), and when COL11A1 was incorporated with the Van Nuys Prognostic Index (B).