The prognostic significance of lysosomal protective protein (Cathepsin A) in breast ductal carcinoma in situ

Michael S Toss[1,2], Islam M Miligy[1,3], Rita Haj-Ahmad[1], Kylie L Gorringe[4,5], Abdulbaqi AlKawaz[1,6], Karuna Mittal[7,] Ian O Ellis[1], Andrew R Green[1] 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, and [7]Georgia 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, proteolytic, CTSA, recurrence

Running title: Cathepsin A expression in DCIS

ABSTRACT

Background: Cathepsin A (CTSA) is a key regulatory enzyme for galactoside metabolism. Additionally, it has a distinct proteolytic activity and plays a role in tumour progression. *CTSA* is differentially expressed at the mRNA level between breast ductal carcinoma *in situ* (DCIS) and invasive breast carcinoma (IBC). In this study, we aimed to characterise CTSA protein expression in DCIS and evaluate its prognostic significance. **Methods:** A large cohort of DCIS (n=776 for pure DCIS and n=239 for DCIS associated with IBC (DCIS/IBC)) prepared as tissue microarray was immunohistochemically stained for CTSA. **Results:** High CTSA expression was observed in 48% of pure DCIS. High expression was associated with features of poor DCIS prognosis including younger age at diagnosis (<50 years), higher nuclear grade, hormone receptor negativity, HER2 positivity, high proliferative index and high hypoxia inducible factor 1 alpha expression. High CTSA expression was associated with shorter recurrence free interval (RFI) (p=0.0001). In multivariate survival analysis for patients treated with breast conserving surgery, CTSA was an independent predictor of shorter RFI (p=0.015). DCIS associated with IBC showed higher CTSA expression than pure DCIS (p=0.04). In the DCIS/IBC cohort, CTSA expression was higher in the invasive component than DCIS component (p<0.0001). **Conclusion:** CTSA is not only associated with aggressive behaviour and poor outcome in DCIS but also a potential marker to predict co-existing invasion in DCIS.

INTRODUCTION

An estimated 20-25% of women undergoing screening mammography are diagnosed with ductal carcinoma *in situ* (DCIS)[1]. DCIS is a non-obligatory precursor to invasive breast carcinoma (IBC) and 45-75% of IBC coexist with DCIS [2-5]. The main objective in treating DCIS is to prevent its progression to invasive disease and to reduce any chance of ipsilateral local recurrence, of which half would be IBC. Achieving this goal is vital, as the mortality rate associated with diagnosis of DCIS alone is only about 1% [6]. Several histopathological characteristics of DCIS can be used as predictors of recurrence including the nuclear grade, presence of comedo necrosis, architectural pattern, lesion size and patient age at diagnosis [7-10]. However, there is a critical requirement for a validated reliable and reproducible clinicopathological and molecular signature that could improve risk stratification of DCIS to facilitate accurate individualised management. Currently available gene signatures such as Oncotype DX DCIS and risk indices for DCIS show lack of reliability and reproducibility between different studies [11-14]. One possible reason is that a considerable percentage of patients (16-25%) are categorised as intermediate risk with undefined further management plan [12, 14] and another possible reason for this lack of reproducibility is the failure to take the tumour microenvironment into account, in view of the advances in deciphering the role of the microenvironment in tumour behaviour [15-22]. Identification an optimal biomarker(s) incorporating tumour intrinsic factors and factors related to the surrounding microenvironment might offer a comprehensive prognostic model in addition to the currently used clinicopathological factors. This in turn would establish the risk of potential progression to IBC and thus identify whether patients required any further treatment, e.g. local re-excision, mastectomy or adjuvant radiotherapy and/or endocrine therapy [23]. In addition, identification of novel markers that predict DCIS invasiveness might also improve our understanding of the disease biology.

Degradation and remodelling of the surrounding basement membrane and stroma are fundamental steps in DCIS progression to invasive disease. The key role of matrix metalloproteinases in stromal degradation is undeniable, however, it is insufficient to

explain DCIS progression to invasive disease depending solely on them. Blocking metalloproteinase action did not show promising results in terms of prevention of disease progression [24, 25].

Cathepsins are a family of lysosomal proteinases or endopeptidases that are highly expressed in numerous human cancers. Lysosomal protective protein/Cathepsin A (CTSA) is an acidic serine carboxypeptidase that has a proactive function towards β-D-galactosidase and N-acetyl-α-neuraminidase [26]. A deficiency in this multifunctional lysosomal protease can cause human lysosomal storage disease (galactosialidosis) [26]. CTSA is expressed in platelets [2], lymphocytes [27] and primary human antigen-presenting cells [28] and plays a role in primary and metastatic human melanocytic tumours [29]. Interestingly, *CTSA* is differentially expressed between IBC and DCIS at the mRNA level [22]. Moreover, while the role of different Cathepsins in breast cancer has been studied before [30, 31], to the best of our knowledge, no previous study has addressed the role of CTSA in DCIS progression and its prognostic impact. In this study, we aimed to evaluate the protein expression of CTSA in a large cohort of DCIS and to assess its prognostic significance.

MATERIAL AND METHODS

Study Cohort

A large DCIS, well characterised annotated cohort, including pure DCIS (n=776) and DCIS coexist with IBC (DCIS-Mixed) (n=239) presented between 1990 to 2012 at Nottingham City Hospital, Nottingham, United Kingdom (UK) was used as previously described [3, 26, 32]. To avoid selection bias, the DCIS-mixed cohort was selected with clinicopathological features comparable to the pure cohort.

Patients' demographic data, method of disease detection either through mammographic screening or symptomatic presentation (regardless of whether the patients had ever been screened), morphological features, treatment including adjuvant radiotherapy (RT) and development of local recurrence were retrieved from local database system. Local recurrence free interval (LRFI) was defined as the time (in months) between 6 months

after the first DCIS surgical removal and occurrence of ipsilateral local recurrence (whatever DCIS or IBC). Patients had close/positive surgical margins or presented with residual tumour tissue and undergoing re-excision surgery within the first 6 months were not considered as recurrence. Patients who developed contralateral disease after initial diagnosis of DCIS were censored at the time of occurrence of the contralateral disease. In the pure DCIS cohort, 83 cases (11%) developed a recurrence within a median follow up period of 103 months (range 6-240), compromising 30 DCIS (36%) and 53 IBC (64%). The majority of the recurrences (n=66) occurred in patients treated with breast conserving surgery (BCS) alone.

In addition, molecular classes, hypoxia inducible factor 1 alpha (HIF-1α) expression and tumour infiltrating lymphocytes (TILs) density were available for the cohort [23, 26, 32]. Briefly, the molecular classes were defined based on the immunohistochemistry (IHC) surrogate classification using oestrogen (ER) and progesterone (PR) receptor, the human epidermal growth factor receptor 2 (HER2) and Ki-67 status. ER and PR were defined as positive if >1% of tumour cells showed nuclear staining [33] while HER2 positivity was defined when more than 10% of tumour cells showed strong complete membranous staining (score +3), where borderline cases (+2) were checked using chromogenic in situ hybridisation technique (CISH) to assess the gene amplification status [34]. Proliferation index was assessed using IHC staining with Ki-67 antibody and defined as low when <14% of cells showed nuclear staining [35]. HIF-1α was evaluated using IHC and was considered positive when >1% of tumour cells showed nuclear staining as previously described [36, 37 3, 26]. Tumours with average number of 20 lymphocytes/duct or more was considered as dense TILs DCIS [23].

Immunohistochemistry

Tissue microarrays (TMAs) were prepared from both cohorts using a TMA GRAND MASTER 2.4-UG-EN MACHINE, using 1 mm punch sets. All cases were reviewed prior to TMA construction and cases with heterogeneous nuclear grade and/or histological

patterns were sampled from all representative areas. A separate TMA from each component of mixed cases (DCIS and IBC) was constructed. The pattern of CTSA expression in malignant breast tissue, adjacent stroma and normal tissue were evaluated using whole tissue sections from 20 cases including 10 pure DCIS and 10 DCIS coexist with IBC.

Primary antibody specificity for rabbit polyclonal CTSA antibody [Ab217857, Abcam, UK] 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). CTSA antibody was used at a dilution of 1:300, which showed a single specific band at the predicted size of 51 KDa (Figure 1A).

Expression of CTSA protein in both DCIS cohorts was assessed by IHC using the Novocastra Novolink TM Polymer Detection Systems kit (Code: RE7280-K, Leica, Biosystems, UK). 4 µm sections were incubated for 24 hours with rabbit polyclonal CTSA (dilution 1:200). Normal kidney tissue was used as a positive control while a negative control was carried out by omitting the primary antibody.

Scoring of CTSA expression

Expression of CTSA in the cytoplasm of tumour cells and surrounding stroma was assessed in each component separately using the semi-quantitative Histo-score (H-score); 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 [38]. Stromal expression was evaluated in stromal fibroblasts, where any staining in the collagenous acellular stroma was not considered. For mixed cohort, each component, DCIS and invasive, was scored separately for the tumor epithelial cells and surrounding stroma. All non-representative cores (lost cores, folded tissue during processing and staining or cores with normal breast tissue or these containing <15% tumour tissue), were excluded from the scoring. Average score was used as a final score for cases with multiple cores (n=180). The cases were scored by two pathologists (MST and IMM) using a multiheaded microscope. For dichotomisation of protein expression, a cut-off point was

defined according to the results from X-tile bioinformatics software (Yale University, version 3.6.1) based on LRFI in the pure DCIS cohort [39]. High CTSA expression was defined as an H-score of >75.

Analysis of *CTSA* mRNA expression in breast cancer:

Due to scarce data on the transcriptomic profiles of DCIS, *CTSA* normalised mRNA expression was evaluated as a potential prognostic marker in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) well characterised IBC cohort dataset (n=1980) [40], to emphasise the prognostic role of CTSA in breast cancer. Moreover, to validate the prognostic significance of CTSA in breast cancer, analysis using the Breast Cancer Gene-Expression Miner v4.1 (bc-GenExMiner v4.1) database was performed.

Statistical analysis

Statistical analyses were performed using SPSS v21 (Chicago, IL, USA) for Windows. Student's t test and analysis of variance (ANOVA) were used to correlate between *CTSA* mRNA level as a continuous variable and other clinicopathological parameters in METABRIC data. Association with *CTSA* mRNA expression and breast cancer specific survival was done after dichotomisation of expression into high and low based on the median value.

Correlation between CTSA expression and other clinical and morphological features in pure DCIS and between pure DCIS and DCIS-mixed cases was performed using Chi-square, unpaired Student's t and ANOVA tests. Wilcoxon signed rank test was used to compare the expression of CTSA between DCIS and invasive components within the DCIS-mixed cohort. Log rank test and Kaplan Meier curves were used for univariate survival analysis against local recurrence free interval. Cox regression model was used for multivariate analysis. 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

Pattern of CTSA expression

The full-face tissue sections demonstrated an even throughout staining of CTSA expression, indicating that TMA is representative for the whole tumour when assessing CTSA expression. Adjacent normal breast terminal duct-lobular units showed weak cytoplasmic staining. Occasional inflammatory cells and stromal fibroblasts were stained in a few cores in the pure DCIS cohort, while most cases in the mixed cohort showed stromal expression in addition to the expression within epithelial tumour cells. When present, CTSA was expressed in the cytoplasm of the epithelial tumour cells. In invasive cases, CTSA was expressed in epithelial tumour cells and surrounding fibroblasts (Figure 1).

After exclusion of uninformative cores, the final number of cases suitable for scoring in was 527 and 204 cores in pure DCIS and DCIS-mixed cohorts, respectively. The median H-score was 75 in pure DCIS (range 0-250); which was similar to the cut-off point generated by X-tile, 90 in the DCIS component of mixed cases (range 0-200), and 110 in the IBC component of the latter (range 0-250). High CTSA expression was observed in 48% of pure DCIS.

The proportion of cases that showed high expression of CTSA was greater in DCIS-mixed than pure DCIS (48% of pure DCIS cases vs. 70% of DCIS with co-existence IBC, χ^2=3.0, p=0.04). Similar results were observed when the score was analysed in a continuous scale (p=0.0001). Moreover, there was a statistically significant difference between CTSA expression within the tumour epithelial cells of the DCIS component and invasive component of the mixed cases (Z=8.8, p<0.0001) (Figure 2).

Significance of CTSA expression in pure DCIS

Numerous clinicopathological parameters indicating poor DCIS prognosis were allied with high CTSA expression (Table 1) including younger age at diagnosis (<50 years), higher nuclear grade, hormonal receptor (ER and PR) negativity, HER2 positivity and high

expression of HIF1α. Moreover, there was a trend of high CTSA expression with symptomatic DCIS presentation, high proliferative (Ki-67) index, and dense TILs. Comparable results were generated when the analysis was carried out using continuous data for CTSA expression (Figure 2).

In the METABRIC cohort, a higher level of *CTSA* mRNA was associated with adverse prognostic parameters including high tumour histological grade (p<0.0001), lymph node metastasis (p=0.001), ER negativity (p<0.0001), HER2 positivity (p<0.0001) (Supplementary Table 1). *CTSA* was also associated with shorter breast cancer specific survival (*HR*=1.5, 95%*CI*=1.1-1.5, p<0.0001) (Supplementary Figure 1). Analysis using the Breast Cancer Gene-Expression Miner v4.1 (bc-GenExMiner v4.1) database showed that high *CTSA* mRNA was associated with higher metastatic relapse (*HR*=1.4, 95%*CI*=1.3-1.6, p<0.0001) (Supplementary Figure 2).

Outcome analysis in pure DCIS cohort

Higher expression of CTSA was correlated with shorter LRFI (all recurrences either as *in situ* or invasive disease) in the entire cohort of pure DCIS (*HR*=3.2, 95%*CI*=1.8-5.8; p=0.0001, Figure 3A). A similar result was seen in patients treated with BCS without any further treatment with adjuvant RT (*HR*=2.7, 95%*CI*=1.6-4.8; p=0.002, Figure 3B). However, there was no correlation between CTSA and outcome in patients treated with either mastectomy or BCS followed by adjuvant RT. Figure 4 displays forest plots (utilising univariate survival analysis) demonstrating the hazard ratio for disease recurrence of the various parameters in patients treated with BCS. In addition, there was a trend between high CTSA expression and occurrence of invasive recurrence (*HR*=1.8, 95%*CI*=0.9-4.6; p=0.07, Figure 3C).

Multivariate survival analysis revealed that high expression of CTSA is an independent poor prognostic factor for tumour recurrence after treatment with BCS (*HR*=2.5, 95% *CI*=1.2-5.3; p=0.015) regardless the other variables including patient age at diagnosis, DCIS size, mode of DCIS presentation, nuclear grade, presence of comedo necrosis, surgical margin width, different molecular classes and adjuvant radiotherapy (Table 2).

Discussion

Understanding the complexity and heterogeneity of DCIS requires full integration of various factors such as tumour-intrinsic factors (e.g. tumour size, nuclear grade, intrinsic molecular subtypes and proliferation rate), tumour-extrinsic tissue factors (e.g. stromal complexity, immune cell response, and the interaction with the surrounding basement membrane and myoepithelium) and clinical factors (e.g. tumour size, grade and patient age at diagnosis). A robust signature combines both clinicopathological and molecular features that incorporates intrinsic tumour factors with the surrounding microenvironment is highly warranted to identify high-risk DCIS, especially those at greatest risk of developing invasive carcinoma [26]. DCIS survival, proliferation and subsequent invasion and metastasis are highly affected by the crosstalk between tumour cells and the surrounding microenvironment including the myoepithelial cell layer, vasculature, stromal fibroblasts and the immune cells. Changes in the breast tumour microenvironment can be detected as early as DCIS or even earlier [2]. Subsequently, epithelial cells and the surrounding microenviroment contribute to reciprocal paracrine acting signalling loops. These actions would stabilise the surrounding immune cells, fibroblasts and myofibroblasts at the DCIS, which in turn alter and remodel the extracellular matrix (ECM) and facilitate tumour cell proliferation, maintenance and invasion [22, 41-43].

Cathepsins are a superfamily of proteins expressed in various types and stages of human cancers [29]. Each member of this family has different functions compared to each other under normal and cancerous conditions. Several cathepsins have been studied and linked to poor prognosis in breast cancer [35]. Similarly, our analysis of the METABRIC cohort showed a correlation between elevated levels of *CTSA* mRNA and aggressive behaviour of IBC. Studies evaluating the role of CTSA in breast cancer and particularly in DCIS are lacking, although a transcriptomic analysis showed *CTSA* is differentially expressed between DCIS and IBC [22]. We were able to validate this result at the protein level by showing that both pure DCIS and the DCIS component of mixed tumours had significantly less CTSA staining than IBC both within the tumour cells and in the

surrounding stromal fibroblasts. Fibroblasts have previously been reported to be an additional source of CTSA activity [34] and our result reflects the interactive role between tumour cells and surrounding tissues in tumour aggressiveness. However, further functional studies are highly recommended to understand the underlying mechanisms and functions of CTSA expression in carcinogenesis and tumour progression either from the tumour cells or the surrounding stroma.

Our analysis of CTSA in a large well characterised cohort of DCIS supported our hypothesis that this protein would be associated with features of high-risk DCIS. In addition, the poor prognostic significance of higher CTSA expression was shown to be independent from other clinical and morphological features, and with a trend of association towards invasive recurrence and progression.

Despite the direct role of CTSA in the ECM degradation that helps in tumour invasion, there are other mechanisms by which CTSA might contribute to tumour aggressiveness. For instance, an increase in CTSA chymotrypsin-like activity was documented to induce the degradation effect of tumour suppressor proteins leading to greater resistance to apoptosis, which would result in more aggressive cancer behaviour [27]. Another mechanism by which CTSA participates in tumour growth is by regulating chaperone-mediated autophagy, which is reported to be essential for cancer cell proliferation [44]. Finally, activation of CTSA requires a low pH acidic environment [34, 45]. Our observation of an association between high CTSA with HIF1α, which usually is accompanied by lower pH, supports elevated CTSA expression as a mechanism of surviving acidic hypoxic environments and more aggressive behaviour. Altogether, these findings support a role in disease progression for the crosstalk and interactions between various factors within the DCIS tumour cells and the surrounding microenvironment.

The role of CTSA in inflammatory processes and in antigen presenting cells function was previously studied [28]. Moreover, overexpression of CTSA in tumour associated macrophages and lymphocytes of the surrounding tissues has been reported [28, 46]. We have previously reported that dense TILs is associated with worse prognosis in DCIS, a

contradictory phenomenon to IBC for which the underlying mechanisms are unclear [23]. Here, we observed a tendency for high CTSA expression in DCIS harbouring dense lymphocytic infiltration that may be associated with an inflammatory function or tumour associated immunity role for CTSA. Further functional mechanistic studies are required to confirm such observation in DCIS and decipher the possible underlying biological mechanisms and significance.

Conclusion

CTSA might have a potential role in DCIS aggressiveness through its regulatory role in ECM degradation and interaction with the surrounding tumour microenvironment. More functional studies to decipher the role of CTSA and its mechanism of action in DCIS behaviour are warranted.

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

REFERENCES

1. Ernster VL, Ballard-Barbash R, Barlow WE *et al.* Detection of ductal carcinoma in situ in women undergoing screening mammography. *Journal of the National Cancer Institute* 2002;**94**;1546-1554.
2. Wong H, Lau S, Yau T, Cheung P, Epstein RJ. Presence of an in situ component is associated with reduced biological aggressiveness of size-matched invasive breast cancer. *British Journal of Cancer* 2010;**102**;1391-1396.
3. Toss MS, Miligy IM, Gorringe KL *et al.* Legumain is an independent predictor for invasive recurrence in breast ductal carcinoma in situ. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2018.
4. Allred DC, Mohsin SK, Fuqua SA. Histological and biological evolution of human premalignant breast disease. *Endocrine-related cancer* 2001;**8**;47-61.
5. Reis-Filho JS, Simpson PT, Gale T, Lakhani SR. The molecular genetics of breast cancer: The contribution of comparative genomic hybridization. *Pathology, research and practice* 2005;**201**;713-725.
6. Kolli N, Garman SC. Proteolytic activation of human cathepsin a. *The Journal of biological chemistry* 2014;**289**;11592-11600.
7. Sweldens C, Peeters S, van Limbergen E *et al.* Local relapse after breast-conserving therapy for ductal carcinoma in situ: A european single-center experience and external validation of the memorial sloan-kettering cancer center dcis nomogram. *Cancer J* 2014;**20**;1-7.
8. Kong I, Narod SA, Taylor C *et al.* Age at diagnosis predicts local recurrence in women treated with breast-conserving surgery and postoperative radiation therapy for ductal carcinoma in situ: A population-based outcomes analysis. *Curr Oncol* 2014;**21**;e96-e104.
9. Cutuli B, Lemanski C, Le Blanc-Onfroy M *et al.* Local recurrence after ductal carcinoma in situ breast conserving treatment. Analysis of 195 cases. *Cancer radiotherapie : journal de la Societe francaise de radiotherapie oncologique* 2013;**17**;196-201.
10. Tunon-de-Lara C, Andre G, Macgrogan G *et al.* Ductal carcinoma in situ of the breast: Influence of age on diagnostic, therapeutic, and prognostic features. Retrospective study of 812 patients. *Annals of surgical oncology* 2011;**18**;1372-1379.
11. Knopfelmacher A, Fox J, Lo Y, Shapiro N, Fineberg S. Correlation of histopathologic features of ductal carcinoma in situ of the breast with the oncotype dx dcis score. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2015;**28**;1167-1173.
12. Rakovitch E, Nofech-Mozes S, Hanna W *et al.* A population-based validation study of the dcis score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast cancer research and treatment* 2015;**152**;389-398.
13. Solin LJ, Gray R, Baehner FL *et al.* A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *Journal of the National Cancer Institute* 2013;**105**;701-710.
14. Yi M, Meric-Bernstam F, Kuerer HM *et al.* Evaluation of a breast cancer nomogram for predicting risk of ipsilateral breast tumor recurrences in patients with ductal carcinoma in situ after local excision. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2012;**30**;600-607.

15. Thompson E, Taube JM, Elwood H *et al.* The immune microenvironment of breast ductal carcinoma in situ. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2016;**29**;249-258.
16. Allen MD, Thomas GJ, Clark S *et al.* Altered microenvironment promotes progression of preinvasive breast cancer: Myoepithelial expression of alphavbeta6 integrin in dcis identifies high-risk patients and predicts recurrence. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2014;**20**;344-357.
17. Van Bockstal M, Lambein K, Gevaert O *et al.* Stromal architecture and periductal decorin are potential prognostic markers for ipsilateral locoregional recurrence in ductal carcinoma in situ of the breast. *Histopathology* 2013;**63**;520-533.
18. Cowell CF, Weigelt B, Sakr RA *et al.* Progression from ductal carcinoma in situ to invasive breast cancer: Revisited. *Mol Oncol* 2013;**7**;859-869.
19. 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 research and treatment* 2012;**133**;1009-1024.
20. Muggerud AA, Hallett M, Johnsen H *et al.* Molecular diversity in ductal carcinoma in situ (dcis) and early invasive breast cancer. *Mol Oncol* 2010;**4**;357-368.
21. Schnitt SJ. The transition from ductal carcinoma in situ to invasive breast cancer: The other side of the coin. *Breast cancer research : BCR* 2009;**11**;101.
22. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast cancer research : BCR* 2009;**11**;R7.
23. Toss MS, Miligy I, Al-Kawaz A *et al.* Prognostic significance of tumor-infiltrating lymphocytes in ductal carcinoma in situ of the breast. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2018;**31**;1226-1236.
24. Gonzalez LO, Gonzalez-Reyes S, Junquera S *et al.* Expression of metalloproteases and their inhibitors by tumor and stromal cells in ductal carcinoma in situ of the breast and their relationship with microinvasive events. *J Cancer Res Clin Oncol* 2010;**136**;1313-1321.
25. Coussens LM, Fingleton B, Matrisian LM. Matrix metalloproteinase inhibitors and cancer: Trials and tribulations. *Science (New York, N.Y.)* 2002;**295**;2387-2392.
26. Toss MS, Miligy IM, Gorringe KL *et al.* Prolyl-4-hydroxylase alpha subunit 2 (p4ha2) expression is a predictor of poor outcome in breast ductal carcinoma in situ (dcis). *British Journal of Cancer* 2018.
27. Hanna WL, Turbov JM, Jackman HL, Tan F, Froelich CJ. Dominant chymotrypsin-like esterase activity in human lymphocyte granules is mediated by the serine carboxypeptidase called cathepsin a-like protective protein. *Journal of immunology (Baltimore, Md. : 1950)* 1994;**153**;4663-4672.
28. Ruszczyk M, Zirpoli G, Kumar S *et al.* Breast cancer risk factor associations differ for pure versus invasive carcinoma with an in situ component in case-control and case-case analyses. *Cancer causes & control : CCC* 2016;**27**;183-198.
29. Kuerer HM. Rational individualised selection of adjuvant therapy for ductal carcinoma in situ. *The Lancet. Oncology* 2011;**12**;2-3.
30. Rochefort H. Cathepsin d in breast cancer: A tissue marker associated with metastasis. *European journal of cancer (Oxford, England : 1990)* 1992;**28A**;1780-1783.

31. Duivenvoorden HM, Rautela J, Edgington-Mitchell LE *et al.* Myoepithelial cell-specific expression of stefin a as a suppressor of early breast cancer invasion. *The Journal of pathology* 2017;**243**;496-509.
32. Miligy IM, Gorringe KL, Toss MS *et al.* Thioredoxin-interacting protein is an independent risk stratifier for breast ductal carcinoma in situ. *Modern Pathology* 2018.
33. 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. *Journal of oncology practice* 2010;**6**;195-197.
34. Rakha EA, Pinder SE, Bartlett JM *et al.* Updated uk recommendations for her2 assessment in breast cancer. *Journal of clinical pathology* 2015;**68**;93-99.
35. 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. *Annals of oncology : official journal of the European Society for Medical Oncology* 2011;**22**;1736-1747.
36. 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**;31-36.
37. van der Groep P, van Diest PJ, Smolders YH *et al.* Hif-1alpha overexpression in ductal carcinoma in situ of the breast in brca1 and brca2 mutation carriers. *PloS one* 2013;**8**;e56055.
38. McCarty KS, Jr., McCarty KS, Sr. Histochemical approaches to steroid receptor analyses. *Seminars in diagnostic pathology* 1984;**1**;297-308.
39. 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**;7252-7259.
40. Curtis C, Shah SP, Chin SF *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;**486**;346-352.
41. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat Med* 2011;**17**;320-329.
42. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* 2011;**144**;646-674.
43. Karnoub AE, Dash AB, Vo AP *et al.* Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* 2007;**449**;557-563.
44. Kon M, Kiffin R, Koga H *et al.* Chaperone-mediated autophagy is required for tumor growth. *Science translational medicine* 2011;**3**;109ra117.
45. Lari SA, Kuerer HM. Biological markers in dcis and risk of breast recurrence: A systematic review. *Journal of Cancer* 2011;**2**;232-261.
46. Jackman HL, Tan F, Schraufnagel D *et al.* Plasma membrane-bound and lysosomal peptidases in human alveolar macrophages. *American journal of respiratory cell and molecular biology* 1995;**13**;196-204.

Table 1: Correlation between CTSA expression with clinicopathological parameters in the pure DCIS cohort

Parameters	CTSA expression in tumour epithelial cells		X^2 (*p*-value)
	Low (N=272) N. (%)	High (N=255) N. (%)	
Age (years)			
≤50	54 (20)	69 (27)	3.8 **(0.045)**
>50	218 (80)	186 (73)	
DCIS Presentation			
Screening	144 (53)	115 (45)	3.2 (0.071)
Symptomatic	128 (47)	140 (55)	
DCIS Size (mm)			
≤20	122 (45)	113 (45)	0.1 (0.998)
>20	150 (55)	139 (55)	
DCIS Nuclear Grade			
Low	44 (16)	23 (9)	10.2 **(0.006)**
Moderate	79 (29)	60 (23)	
High	149 (55)	172 (68)	
Comedo necrosis			
Yes	174 (64)	177 (69)	1.8 (0.18)
No	98 (36)	78 (31)	
Oestrogen receptor (ER) status			
Negative	49 (20)	79 (34)	11.9 **(0.001)**
Positive	199 (80)	149 (66)	
Progesterone receptor (PR) status			
Negative	89 (44)	112 (56)	6.4 **(0.011)**
Positive	152 (56)	119 (44)	
HER2 status			
Negative	219 (85)	169 (70)	17.2 **(<0.0001)**
Positive	37 (15)	72 (30)	
Proliferation Index (Ki-67)			
High (≥14%)	41 (20)	58 (28)	3.7 (0.051)
Low (<14%)	168 (80)	152 (72)	
Molecular classes			
Luminal A	113 (59)	89 (43)	11.6 **(0.009)**
Luminal B	33 (17)	42 (20)	
HER2 enriched	19 (10)	40 (20)	
Triple negative	28 (14)	37 (17)	
Tumour infiltrating lymphocytes (TILs)			
Dense	90 (44)	102 (53)	3.1 (0.075)
Sparse	116 (56)	92 (47)	
Hypoxia inducible factor 1 alpha (HIF1α)			
High	35 (18)	60 (32)	9.4 **(0.002)**
Low	159 (82)	130 (68)	

Significant *p* values are in bold

Table 2: Multivariate survival analysis (Cox regression model) of variables predicting outcome in terms of ipsilateral local recurrence in patients treated by breast conserving surgery

Parameters	Hazard ratio (*HR*)	95.0% *confidence interval (CI)*		Significance *p*-value
		Lower	Upper	
High CTSA expression	2.5	1.2	5.3	**0.015**
Patient Age	0.6	0.3	1.4	0.255
DCIS size	1.1	0.5	2.2	0.811
DCIS nuclear Grade	1.7	0.9	3.1	0.056
Comedo type necrosis	0.6	0.3	1.4	0.254
Radiotherapy	0.2	0.1	0.8	**0.019**
Surgical margins	0.8	0.4	1.4	0.381
Molecular classes	0.9	0.7	1.3	0.7
DCIS presentation	2.1	0.9	4.1	0.051

*This model includes CTSA expression in tumour cells with other known parameters determining aggressive DCIS behaviour in the pure DCIS cohort.
***p* values in bold= Significant**

Supplementary Table 1: Correlation between *CTSA* mRNA level and the clinicopathologic parameters in the METABRIC series of invasive breast cancers (n=1980).

Parameter	Number of cases	Mean *CTSA* mRNA level	*p*-value
Patient Age (years)			
<50	383	8.5	0.452
≥50	1556	8.5	
Tumour Size (mm)			
≤20	622	8.4	0.007
>20	1331	8.5	
Histologic Grade			
1	170	8.3	
2	770	8.4	**<0.0001**
3	952	8.6	
Lymph node metastasis			
Negative	1035	8.4	**0.001**
Positive	938	8.5	
Oestrogen Receptor Status			
Positive	1506	8.4	**<0.0001**
Negative	474	8.6	
HER2 Status			
Negative	1733	8.4	**<0.0001**
Positive	247	8.6	
PAM50 molecular classes			
Luminal A	718	8.4	
Luminal B	488	8.6	**<0.0001**
Basal-like	329	8.6	
HER2 enriched	240	8.7	
Normal like	199	8.5	

***p* values in bold= significant**
METABRIC: Molecular Taxonomy of Breast Cancer International Consortium

Figures

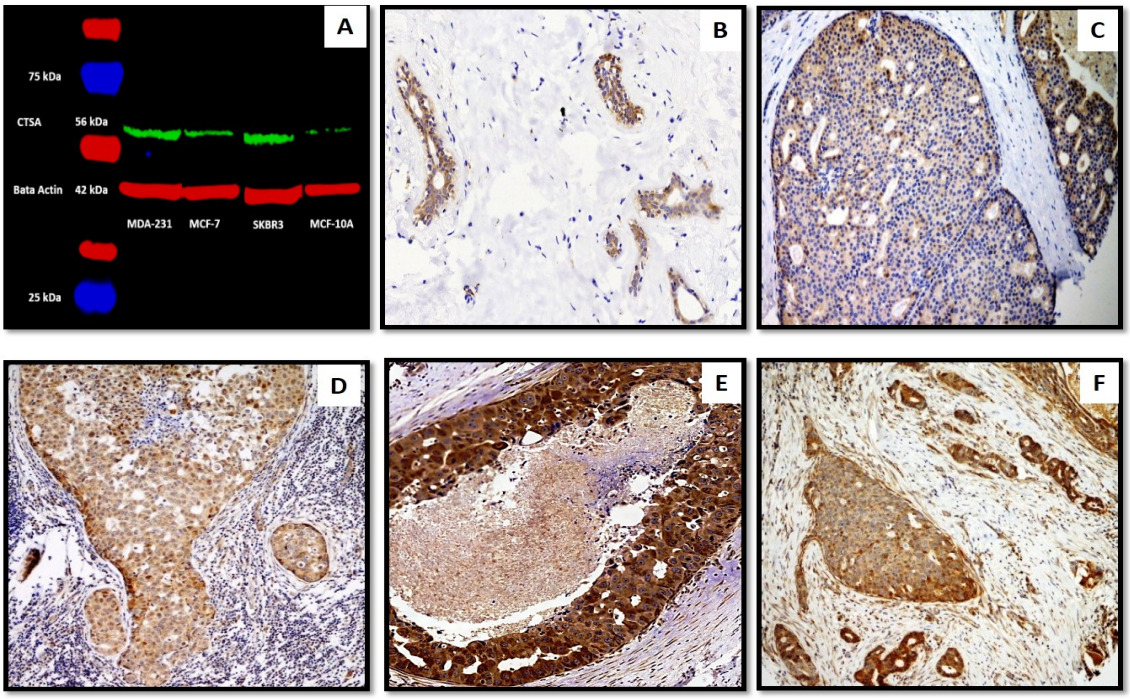


Figure 1: Anti-CTSA antibody validation and patterns of protein expression. A) Western blot of rabbit polyclonal anti-CTSA antibody showing a single specific band (upper green band) at expected molecular weight (51 kDa) in MDA-MB-231, MCF-7, SKBR3 and MCF-10A cell lysates. The lower red band represents the beta-actin (positive control) at 42kDa molecular weight, B) Normal breast duct (x20) shows weak cytoplasmic staining of CTSA in the normal epithelial cells. C) Weak CTSA expression (x40) in a pure low grade DCIS case; D) stronger expression of CTSA in (x40) in a pure DCIS case. Note the dense inflammatory cells surrounding the DCIS. E) Strongest expression of CTSA (x40) in a pure high grade DCIS case. F) Expression of CTSA in a mixed case (x20) showing higher intensity within the invasive tumour cells than the DCIS component. Stromal expression of CTSA is also noticed.

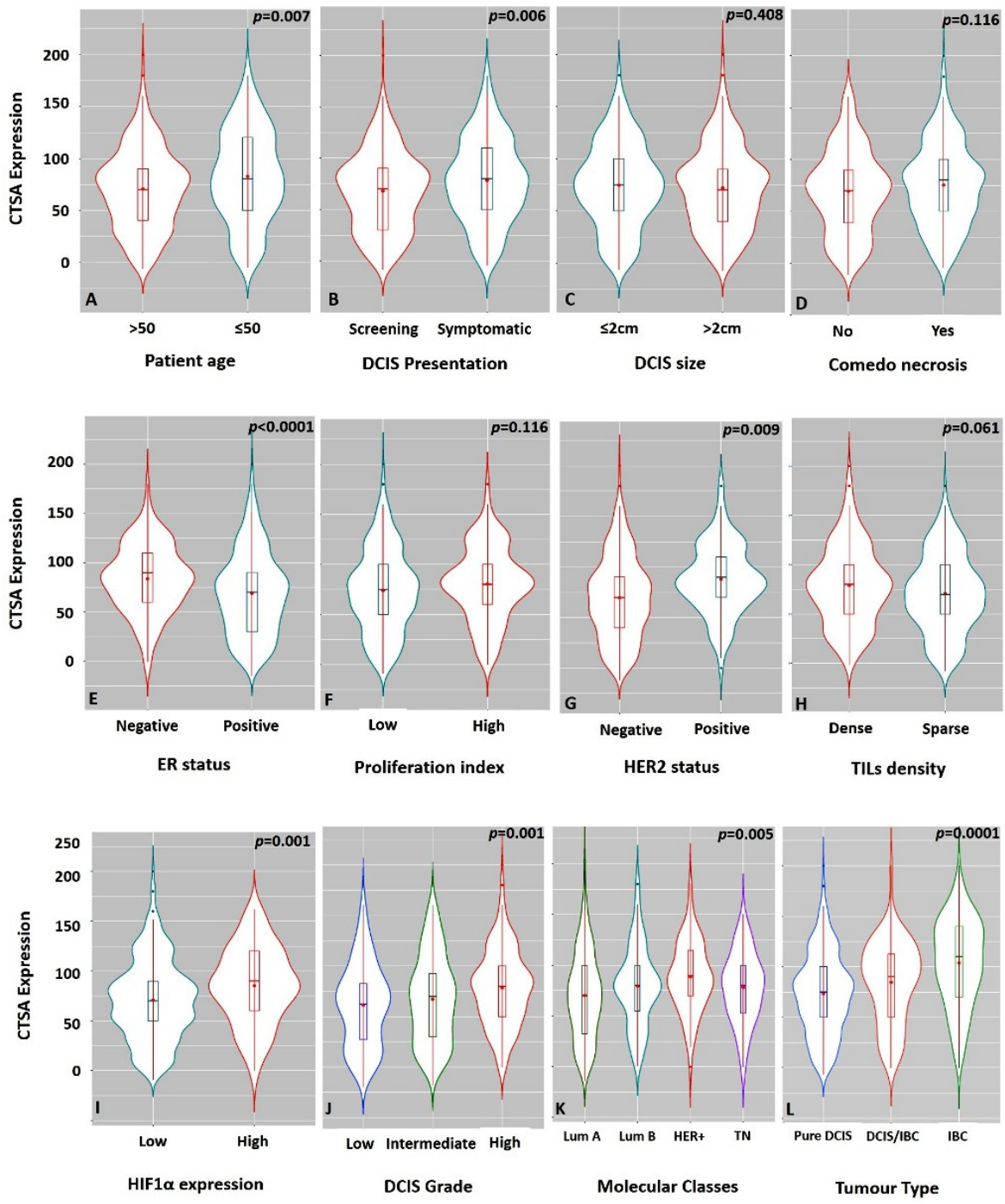


Figure 2: Violin plots showing the associations between CTSA expression (H score) and other clinicopathological parameters in pure DCIS cohort (A-K) and the difference between CTSA expression in pure DCIS and DCIS mixed with invasive carcinoma (L), using continuous expression data. The central boxplot represents 95% confidence interval and the median while the red dot represents the mean. (TILs; tumour infiltrating lymphocytes, HIF1α; hypoxia inducible factor 1 alpha, Lum; luminal, TN; triple negative, DCIS; ductal carcinoma in situ, IBC; invasive breast cancer)

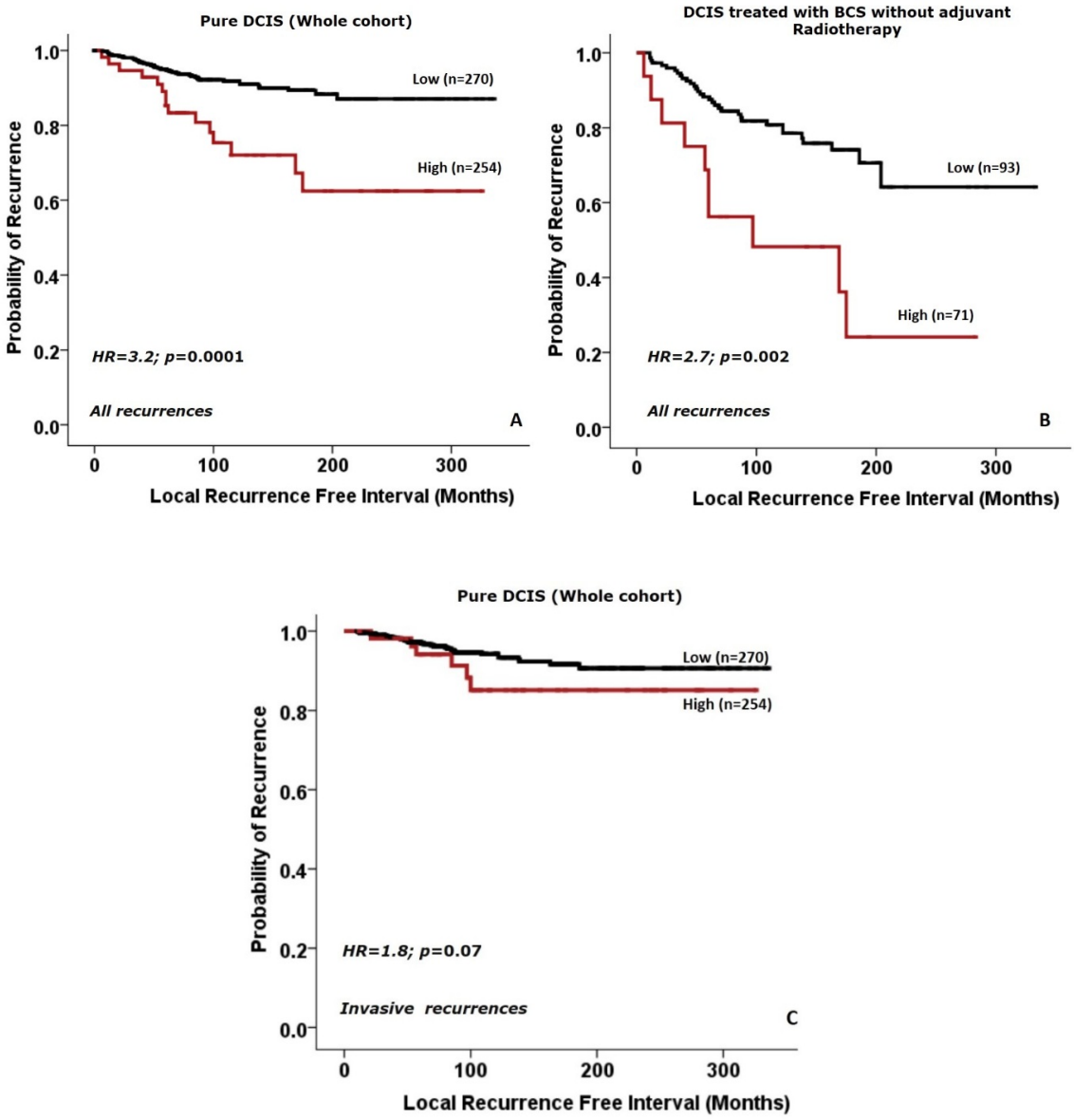


Figure 3: Kaplan Meier curves show that high expression of CTSA is associated with shorter ipsilateral local recurrence free survival (LRFS) in the whole series (A), and in patients treated with breast conserving surgery (BCS) without adjuvant radiotherapy (B). (C) showing the association between CTSA expression and invasive recurrences only in the whole cohort.

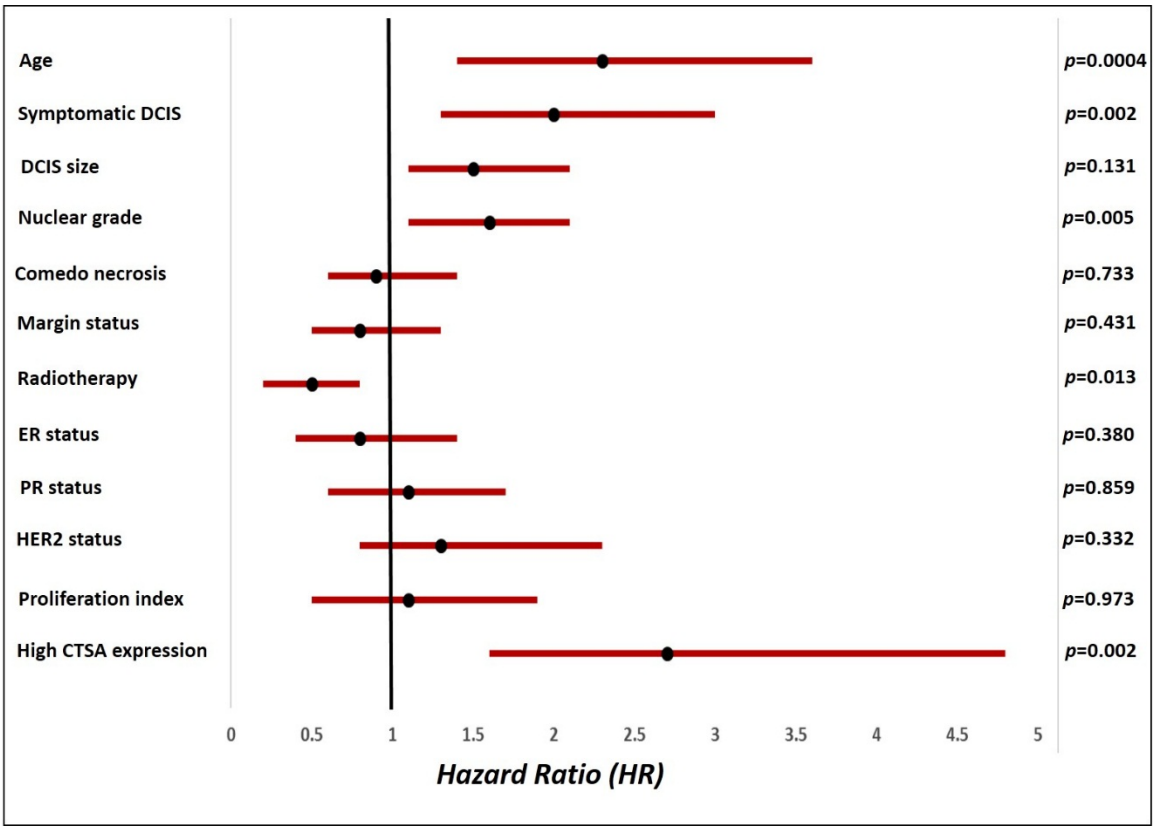
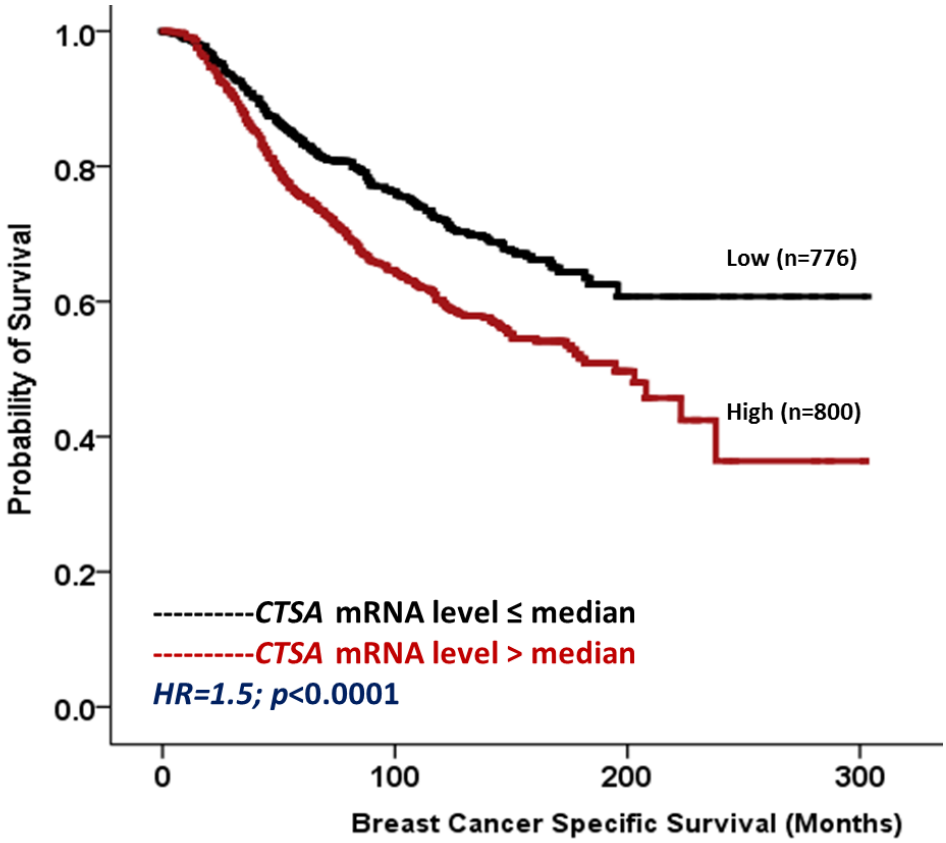
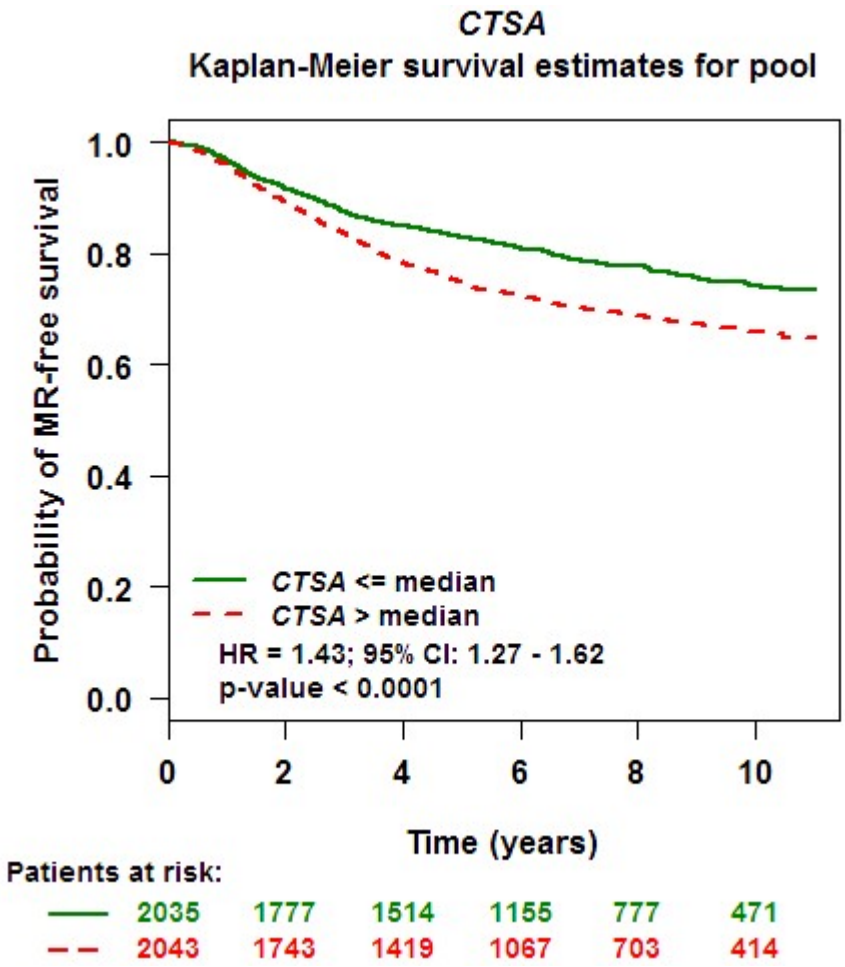


Figure 4: Forest plot showing the univariate analysis results of association between different clinicopathological parameters including CTSA expression and ipsilateral tumour recurrence for patients treated with breast conserving surgery in the pure DCIS cohort

Supplementary figure



Supplementary Figure 1: Association between *CTSA* 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 based on the median.



Supplementary Figure 2: Association between *CTSA* mRNA level and metastatic free interval in Breast Cancer Gene Miner Data. The cohort was split into high and low mRNA expression based on the median.