

Mediator complex (MED) 7: a biomarker associated with good prognosis in invasive breast cancer especially ER+ luminal subtypes

Chitra Joseph^a, Olivia Macnamara^a, Madeleine Craze^a, Roslin Russell^b, Elena Provenzano^c, Christopher C. Nolan^a, Maria Diez-Rodriguez^a, Sultan N Sonbul^a, Mohammed A Aleskandarany^a, Andrew R. Green^a, Emad A. Rakha^a, Ian O. Ellis^a, Abhik Mukherjee^a

^a Division of Cancer and Stem Cells, School of Medicine, University of Nottingham and Nottingham University Hospitals NHS Trust, City Hospital Campus, Nottingham NG5 1PB, UK.

^bCRUK Cambridge Research Institute, Cambridge, United Kingdom;

^cAddenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust, Cambridge, United Kingdom.

Correspondence:

Dr Abhik Mukherjee. Academic Pathology, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham City Hospital, Hucknall Road, Nottingham NG5 1PB
Tel: (44) 115 8231407, Email: Abhik.Mukherjee1@nottingham.ac.uk

Running title: MED7 in breast cancer

Conflict of interest: The authors confirm that they do not have any conflict of interests to declare.

ABSTRACT

Background: Mediator complex (MED) proteins have a key role in transcriptional regulation, some interacting with the oestrogen receptor (ER). Interrogation of the METABRIC cohort suggested that MED7 may regulate lymphovascular invasion (LVI). Thus, MED7 expression was assessed in large breast cancer (BC) cohorts to determine clinico-pathological significance.

Methods: MED7 gene expression was investigated in the METABRIC cohort (n=1980) and externally validated using bc-GenExMiner v4.0. Immunohistochemical expression was assessed in the Nottingham primary BC series (n=1280). Associations with clinicopathological variables and patient outcome were evaluated.

Results: High MED7 mRNA and protein expression was associated with good prognostic factors: low grade, smaller tumour size, good NPI, positive hormone receptor status ($p < 0.001$), and negative LVI ($p = 0.04$) status. Higher MED7 protein expression was associated with improved BC specific survival within the whole cohort and ER+/luminal subgroup. Pooled MED7 gene expression data in the external validation cohort confirmed association with better survival, corroborating with the protein expression. On multivariate analysis, MED7 protein was independently predictive of longer BC specific survival in the whole cohort and Luminal A subtype ($p < 0.001$).

Conclusions: MED7 is an important prognostic marker in BC, particularly in ER+ luminal subtypes, associated with improved survival and warrants future functional analysis.

Keywords: MED7, Breast Cancer, METABRIC, LVI, Luminal, ER+

1. INTRODCUTION

Breast cancer (BC) is a heterogeneous disease varying in presentation, morphological types, response to therapy and patient outcome. The development of high-throughput technologies to investigate genetic, epigenetic and proteomic changes has helped to unravel the complexity of BC biology (Dawson *et al*, 2013). Metastasis is the major cause of morbidity and mortality in BC patients and lymphovascular invasion (LVI), rather than blood vascular invasion seems to be the major mechanism involved in the early stages of metastases in BC (Mohammed *et al*, 2007). LVI has been shown to correlate with an increased histological grade and a poor prognosis. (Rakha *et al*, 2012). The molecular mechanisms of LVI are complex, involving multiple pathways, and remain hugely unknown (Karaman & Detmar, 2014). The METABRIC study investigated the genomic and transcriptomic data across 2,000 breast tumours, where associations between germline variants (CNVs and SNPs), somatic aberrations (CNAs) and alterations in gene expression were found. Further clustering analysis have identified 10 novel clusters (IntClust 1-10) which were associated with distinct CNAs and gene expression changes. The clusters have further divided the molecular subtypes and were associated with different clinical outcomes and drivers. In this study, a novel approach was utilised to unravel the molecular determinants of LVI. Global gene expression profiling data from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (Curtis *et al*, 2012) was utilised to identify differentially expressed genes related to LVI status as determined by central histopathology review of H&E sections supplemented by D2-40 immunohistochemical analysis. Among top differentially expressed genes, the mediator protein MED7 was found to be inversely associated with LVI.

The Mediator complex is a multi-subunit transcriptional coactivator complex required for the transcription of nearly all protein-coding genes (Gustafsson & Samuelsson, 2001). By its direct association with both signal-activated transcription factors and the RNA polymerase II transcription machinery, the mediator proteins function as a general integrator of regulatory signals that congregate on the promoters of protein-coding genes (Woychik & Hampsey, 2002). Disruption to Mediator subunits can affect many different cellular functions and fates, some of which could potentially be involved in carcinogenesis (Schiano *et al*, 2014). One of the Mediator subunits; MED7, plays an

important role in gonadal development and embryogenesis (Kwon *et al*, 1999). However, despite being a highly-conserved Mediator subunit, MED7 is not essential for viability across all species. A study investigating the functions of MED7 in *Candida albicans* found that a loss of MED7 did not lead to loss of viability (Tebbji *et al*, 2014). However, loss of MED7 has been reported to have a substantial impact on several cellular functions, in particular impairing metabolic functions. Overall, by deletion of different Mediator sub-modules, including MED7N (the N-terminal subunit of MED7) metabolic sensing, stress response, and certain amino-acid biosynthesis pathways are affected (Koschubs *et al*, 2009). Gene expression profiling in gastrointestinal stromal tumours has shown that MED7 down-regulation is associated with an increased tumour risk and could therefore potentially be a marker of favourable prognosis (Hur *et al*, 2010). Other members of the mediator complex such as MED1 have been reported to be associated with LVI in other series (Fidalgo *et al*, 2015) and hence MED7, another member of the Mediator complex, was deemed to be an interesting candidate to investigate. MED1 and MED24 have also been found to interact to mediate oestrogen receptor (ER) functions and regulate pubertal mammary gland and BC development (Hasegawa *et al*, 2012). Because of the association of other mediator proteins with ER, the role of MED7 in ER+ BCs was also explored, as ER+ BCs remain the most heterogeneous molecular group (Ciriello *et al*, 2013; Stingl & Caldas, 2007). Overall, this study aimed to investigate the clinico-pathological and biological significance of MED7 in BC including its role in LVI and hormonal receptor status.

2. MATERIALS AND METHODS

2.1 Differential gene expression and the selection of *MED7*:

LVI status, as defined by morphology (H&E) supplemented with immunohistochemistry (D2-40 and CD34), was available for the Nottingham subset of the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort and hence interrogated for differences between LVI+ and LVI- subgroups. Lymph node-positive samples were excluded from LVI- subgroup to avoid the confounding effect of undetectable LVI in these tumours. Data on differentially expressed genes between LVI+ and LVI- subgroups was obtained from microarray analysis data normalised to fit into a linear regression model (LIMMA: Linear models of microarray analysis). *MED7* was ranked within the top 10% of these candidate genes as inversely (log fold change: -1.29) correlated with LVI (adjusted p value =0.0005). Chosen as a gene of interest, data from the whole (METABRIC) cohort (Supplementary Table 1; n=1980) (Curtis *et al*, 2012) was used to evaluate mRNA expression of *MED7*. [The METABRIC study provides data on genomic and transcriptomic profiling of breast cancer using the Affymetrix SNP 6.0 and Illumina HT-12 v3 platforms respectively. Detailed description of the experimental assays and analytical methods used have been described previously (Curtis *et al*, 2012; Silwal-Pandit *et al*, 2014)]. The assessment of the clinico-pathological impact of *MED7* transcription and its associations with clinical outcome in the whole METABRIC series was performed by setting a cut-off point for the mRNA expression of *MED7* at the median.

2.2 External validation cohort

For external validation, *MED7* mRNA expression was interrogated through the BC-GenExMiner v4.0 (Breast Cancer Gene-Expression Miner v4.0 online dataset (<http://bcgenex.centregauducheau.fr>) (Jezequel *et al*, 2012), also used in other published studies (Aleskandarany *et al*, 2016a). This is composed of two statistical mining modules; the ‘prognostic module’, offering the possibility to evaluate the *in vivo* prognostic impact of candidate genes in BC, and the ‘correlation module’, to compute correlation coefficients between gene expressions or to find lists of correlated genes in BC. In this external validation, the prognostic module of the BC-GenExMiner, that evaluates the *in vivo*

prognostic impact of candidate genes in BC was utilised, using Cox model and Kaplan–Meier plot generation (Jezequel *et al*, 2012). Datasets with available MED7 expression in this online repository are catalogued in Supplementary Table 2.

2.3 Patients and tumours

The well-characterised Nottingham Tenovus Primary BC Series comprised the study population for protein expression (Supplementary Table 3). Briefly, this comprised of women aged 70 or less who presented to Nottingham City Hospital from 1988-1998 and received uniform treatment with a long-term follow-up period. Patients' clinicopathological profiles included histological phenotype, molecular subtypes, primary tumour size, histological grade, tumour stage, nodal status, Nottingham Prognostic Index (NPI), receptor status and other BC related biomarkers (Abd El-Rehim *et al*, 2005; Rakha *et al*, 2009). Outcome data including breast cancer-specific survival (BCSS), distant metastasis-free interval (MFI) was maintained on a prospective basis. BCSS was defined as the interval (in months) from the date of primary surgery to the time of death because of BC, while MFI identified as the interval from the date of primary surgery to the appearance of distant metastasis.

2.4 Western blotting

For validation of MED7 Rabbit monoclonal antibody [EPR15410 (Abcam- Ab187146, Cambridge, UK) specificity, Western blotting was performed on whole cell lysates of MCF-7, SKBr3 and HEK293 (obtained from the American Type Culture Collection; Rockville, MD, USA) cell lines using 1:1000 dilution of the primary antibody, and fluorescent secondary antibodies (1:15000) (IR Dye 800CW donkey anti-rabbit and 680RD donkey anti-mouse, LI-COR Biosciences, UK). 5% milk (Marvel Original Dried Skimmed Milk, Premier Food Groups Ltd, St Albans, UK) was used for blocking. Mouse β -Actin (A5441, Sigma-Aldrich; Clone AC-15; Sigma, UK) at 1:5000 was used as a house-keeping protein. A protein ladder (Page Ruler Plus Prestained Protein Ladder, ThermoScientific, Waltham, MA, USA) was included. To visualise bands, fluorescence at wavelengths of 600, 700 and 800nm was used on a Licor Odyssey Fc with image studio 4.0 (LI-COR Biosciences).

2.5 Protein expression by Immunohistochemistry

Tumour samples were arrayed onto tissue microarrays (TMAs) as previously described (Abd El-Rehim *et al*, 2005). IHC was performed on TMA sections using the Novolink Max Polymer Detection system (Leica, Newcastle, UK). In brief, sections were deparaffinised with xylene and rehydrated through 100 % ethanol. Heat-induced retrieval of antigen epitopes was performed in citrate solution (pH 6.0). MED7 staining was performed with a rabbit monoclonal antibody [EPR15410 (Abcam- Ab187146, Cambridge, UK)], diluted (1:50), incubated for 60 min at room temperature. 3-3' Diaminobenzidine tetrahydrochloride (Novolink DAB substrate buffer plus) was freshly prepared and used as chromogen. Counter staining was performed using Meyer's haematoxylin for 6 min. Negative (omission of the primary antibody) and positive controls (anti-human- β -2-microglobulin; A0072, Dako) were included.

The modified H-score method was used in assessing IHC staining, taking the staining intensity and percentage positivity into account (McCarty & McCarty, 1984). Briefly the percentages of positively stained tumour cells for each of these intensities were subjectively estimated. Staining intensity (0–3) was multiplied by percentage (0–100) and final scores were obtained, giving a range of 0 to 300. High-resolution digital images were generated via scanning the IHC stained slides (Nanozoomer; Hamamatsu Photonics, Welwyn Garden City, UK) at x20 magnification to facilitate the scoring of the TMA cores using a high-resolution (1920x1080) screen. Staining was double scored blindly by two researchers including a consultant histopathologist for ~25% cores to assess inter-observer concordance. Immunohistochemical staining and dichotomisation of the other biomarkers included in this study were as per previous publications (Aleskandarany *et al*, 2012; Aleskandarany *et al*, 2010a; Aleskandarany *et al*, 2011; Aleskandarany *et al*, 2010b; Barros *et al*, 2014; Elsheikh *et al*, 2008; Habashy *et al*, 2013; Rakha *et al*, 2007; Rakha *et al*, 2009) (Supplementary Table 4). BC molecular subtypes were defined based on the IHC profile as: Luminal A: ER+/HER2- Low Proliferation (Ki67<10%), Luminal B: ER+/HER2- High Proliferation (Ki67 \geq 10%), HER2-positive class: HER2+ regardless of ER status, Triple Negative (TPN): ER-, PgR- and HER2-.

2.6 Statistical analysis

IBM SPSS 22.0 (Chicago, IL, USA) software was used for statistical analysis. Univariate analysis was performed using the chi-squared test to evaluate the significance of the association between expression of the biomarkers and the clinicopathological parameters of the data, as well as other previously investigated biomarkers. Kaplan–Meier analysis with a log-rank test for significance was performed to assess BCSS and DFI survival differences. Multivariate Cox Regression analysis with adjustment of co-variables was fitted to test independence from standard prognostic factors in BC (stage, grade and LVI). A p value of < 0.05 was considered significant.

2.7 Ethics

This study was approved by the Nottingham Research Ethics Committee 2 under the title ‘Development of a molecular genetic classification of breast cancer’. All samples from Nottingham used in this study were pseudo-anonymised and collected prior to 2006 and therefore under the Human Tissue Act informed patient consent was not needed. Release of data was also pseudo-anonymised as per Human Tissue Act regulations.

3 RESULTS

3.1 MED7 mRNA expression and clinicopathological parameters:

In the METABRIC cohort, high *MED7* mRNA expression was observed in 981 cases (49.6%). High *MED7* expression was associated with lower grade ($p < 0.0001$), older age ($p < 0.0001$) and good/excellent prognostic histological types (lobular and mucinous types ($p = 0.001$)). It was also associated with ER+ and progesterone receptor positive (PR+) tumours (both $p < 0.0001$) and HER2-negative status ($p = 0.00001$; Table 1). When comparing the levels of *MED7* mRNA expression in the intrinsic (PAM50) subtypes, significant correlations were observed with luminal subtype tumours while the basal subtype showed the least expression levels ($p < 0.0001$; Figure 1A). High *MED7* mRNA expression was significantly associated with IntClusts 3 and 8 ($p < 0.00001$), clusters known to be enriched for luminal A like BCs associated with the most favourable clinical outcome in the METABRIC cohort. Overexpression of *MED7* mRNA displayed significantly improved patients' survival in the whole cohort (Figure 1B; $p = 0.025$). There was no association between *MED7* mRNA expression and outcome in any of the PAM50 subtypes (Supplementary Figure 1A-D). As stated in the materials and methods, *MED7* was inversely (log fold change: -1.29) correlated with LVI (adjusted p value = 0.0005) when interrogated on the Nottingham subset of the cohort, that had available LVI data.

3.2 MED7 expression and BC-biomarkers

For IHC analysis on the Nottingham BC series, the specificity of the antibody was validated with a single specific band at the predicted size (32 kDa; Figure 1C). *MED7* IHC showed nuclear staining with no cytoplasmic or stromal staining (Figure 1D-F). The H-Scores of *MED7* nuclear expression did not follow a normal distribution and hence the cut-off point for the *MED7* H-score for low/high was set by the SPSS programme at the median (H score > 130). All cut-offs were set before analysis. Of the 1280 informative cores, 637 (49.8%) had high *MED7* expression and 643 (50.2%) showed low expression. Similar to the mRNA observation, protein expression was also associated with good prognostic parameters. High nuclear *MED7* expression was associated with smaller tumour size ($p < 0.0001$), lower

grade ($p < 0.0001$), lower mitotic scores ($p < 0.0001$), higher tubule formation ($p = 0.0004$) and less nuclear pleomorphism ($p < 0.0001$). Lobular carcinomas showed significantly higher expression of MED7 ($p = 0.001$) in comparison to ductal no special type and medullary subtypes (Table 2). Loss of MED7 protein was correlated with positive LVI status ($p = 0.04$). Overall, a relative lack of MED7 correlated with a poorer NPI ($p < 0.0001$).

On IHC, high nuclear MED7 expression showed significant positive association with ER/PR status ($p < 0.0001$), while negative association was observed with basal cytokeratins CK5/6 ($p < 0.001$) and CK17 ($p = 0.003$). Basal-like breast cancer is highly heterogeneous associated with high grade, poor patient outcome and CK5/6, and CK17 expression. MED7 expression was correlated with other characterized biomarkers on the series, some explored for their known association with ER+ luminal subtypes, viz coactivator-associated arginine methyltransferase 1, CARM1, an ER α coactivator (Al-Dhaheer *et al*, 2011), the ER-chromatin interaction regulator Forkhead box protein A1, FOXA1 (Hurtado *et al*, 2011) and RAS-like estrogen regulated growth inhibitor, RERG (Habashy *et al*, 2011). High expression of MED7 was positively associated with these luminal subtype related bio-markers: CARM1 ($p = 0.015$), RERG ($p = 0.015$) and FOXA1 ($p < 0.00001$). Positive correlations were observed with cell cycle regulatory proteins such as GATA3 ($p < 0.0001$), and STAT3 ($p < 0.0001$), markers also known to be highly expressed in ER+ BC associated with favourable outcome (Aleskandarany *et al*, 2016b; Voduc *et al*, 2008). Expression of histone modifiers which influence hormone responsive gene expression in BC (Nagarajan *et al*, 2015) were also positively associated with MED7 expression: viz. histone methylation modifiers at lysine (H3K4Me2: $p = 0.041$; H4K12ac: $p = 0.004$) and arginine residues (H4R3Me2; $p = 0.048$) (Table 3). In contrast, negative correlations were observed with proliferation markers such as Ki67 ($p = 0.002$), epithelial-mesenchymal transition markers such as N-cadherin ($p < 0.0001$) and signaling pathway biomarkers like phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (PIK3CA; $p = 0.001$) and the epidermal growth factor receptor (EGFR; $p = 0.009$).

3.3 MED7 and association with patient outcome

High expression of MED7 protein was predictive of longer BCSS in the whole cohort ($p < 0.0001$, HR=0.66; 95%CI: 0.54-0.81; Figure 2A), Luminal A ($p = 0.009$, HR=0.42; 95%CI: 0.21-0.82; Figure 2B) and Luminal B ($p = 0.014$, HR=0.66; 95%CI: 0.47-0.92; Figure 2C) subtypes. There was no association between MED7 protein and outcome in TPN ($p = 0.442$, HR=0.83; 95% CI: 0.51-1.3; Figure 2D) and Her2+ subgroups ($p = 0.885$, HR=1.0; 95% CI: 0.66-1.62; Figure 2E). However, MED7 expression was not significantly associated with MFI ($p = 0.056$, HR=0.84; 95% CI: 0.70-1.0; Figure 2F). In multivariate Cox regression analysis, MED7 protein was a predictor of better BCSS in the whole cohort and Luminal A subtypes. ($p < 0.01$, Table 4).

3.4 The prognostic impact of MED7 mRNA expression using bc-GenExMiner v 4.0 (Breast Cancer Gene-Expression Miner v 4.0).

Targeted prognostic analyses for *MED7* in LN negative BCs ($n = 34$ data sets, 2431 patients) indicated that high levels of gene expression correlated with adverse event free survival ($p = 0.0085$; HR=0.82; 95% CI: 0.71-0.95, Figure 2G). As LVI status was not available for these cohorts, adverse event free survival served as a surrogate for the early metastatic phenotype. Expression analysis for *MED7* with respect to hormone status (ER, PgR and HER2; $n = 5461$; Supplementary Figure 2A-C) corroborates with our analyses on the METABRIC/Nottingham cohorts that higher MED7 correlated with ER/PR positive status ($p < 0.0001$) and but not HER2 status ($p > 0.05$). Similarly, interrogation of the BC gene miner data revealed correlations with lower grade ($n = 3421$; $p < 0.0001$) and better NPI ($n = 1684$; $p = 0.0021$).

4. DISCUSSION

Despite the plethora of biomarkers studied in BC, the molecular signatures underlying LVI in BC is undefined. Differential gene expression analysis in strictly defined LVI strata (morphology supplemented by IHC) in well validated BC cohorts would potentially provide the ideal opportunity to interrogate key molecules involved in generating the early metastatic phenotype. Using this approach utilising the METABRIC dataset, the mediator subunit MED7 was identified to be negatively related to LVI. To date, there has been relatively little research into MED7 compared with other Mediator subunits with respect to its potential role in BC. Of other mediator subunits, MED1 is required for estrogen receptor-mediated gene transcription and breast cancer cell growth (Zhang *et al*, 2013) and has also been implicated in vascular invasion (Fidalgo *et al*, 2015). On the other hand, recurrent MED12 somatic mutations have been seen in fibroadenomas and phyllodes tumours (Lien *et al*, 2016). In contrast, the functional relationships of MED7 are little known. The aim of the study was to assess the potential associations between MED7 protein and various clinico-pathological variables in primary BC including hormonal receptor status and LVI to investigate its potential as a prognostic tool.

High expression of both MED7 mRNA and protein were significantly associated with better behaving tumour characteristics, viz. low histological grade, older age, good NPI, ER+/PR+ tumours and histological subtypes of good prognosis. This is comparable to studies in other tumours where MED7 down-regulation was significantly associated with increased risk of gastrointestinal stromal tumours (Koschubs *et al*, 2009). MED7 was preferentially positive in the lobular carcinomas in contrast to ductal or medullary-like tumours. Invasive lobular carcinomas are classically of lower grade with a low rate of mitosis and relatively uniform nuclei (Haltas *et al*, 2012; Rakha *et al*, 2008). Many classical-type ILC tumours express ER and PR (Yu *et al*, 2011) both of which were significantly positively associated with MED7 expression. It was hypothesised that low-grade BCs develop through a different pathway from high-grade tumours which may involve alterations in expression of ER and altered genetic profiles in low-grade compared with high-grade tumours (Abdel-Fatah *et al*, 2008). This could implicate MED7 as having a putative role in low-grade ER+ tumourigenic pathways. Moreover, high MED7 mRNA

expression level was significantly associated with the ER positive integrative clusters 3 and 8 which had the most favourable clinical outcome in the METABRIC study (Curtis *et al*, 2012).

There are several pathways by which ER is able to activate gene transcription. In the ER mediated pathway, dimerised ER directly binds to DNA sequences called Estrogen Response Elements (EREs) in relevant activated genes. However, ER is also known to use non-classical pathways to activate these genes either via Activator protein 1 (AP-1) or via Specificity protein 1 (Sp-1). MED7 acts as a co-regulator for Sp1 activity (Ryu *et al*, 1999) and it is therefore possible that MED7 acts within the non-classic Sp-1 pathway of ER gene activation. The biomarkers characterized on the Nottingham Primary series also serve as indicators of possible molecular networks in ER+ tumours where MED7 may be an interacting partner. Markers known to be overexpressed in luminal BC viz, CARM1 (Al-Dhaheri *et al*, 2011), RERG (Finlin *et al*, 2001) and FOXA1 (Laganiere *et al*, 2005) revealed significant positive association with MED7 as also luminal CKs, steroid receptors, and cell cycle inhibitors (p21 and p27) which are associated with good prognostic characteristics. MED7's positive association with luminal markers indicate its role in better behaving tumours. However, the role of MED7 within the ER-related pathways may be quite complex, depending on the specific interacting partner. For example, in this study, MED7 expression was found to be negatively associated with EGFR expression. On one hand, it is known that EGFR overexpression in BC is associated with increased tumour size and worse patient outcomes and negatively correlates with estrogen receptor status (Masuda *et al*, 2012), explaining the observed negative association with MED7. However, it is also known that activation of EGFR by EGF triggers phosphorylation of MAPK and ERK, which in turn causes phosphorylation of Ser118 of the AF-1 domain of ER α , resulting in ER transactivation (ElTanani & Green, 1997). This transactivation can also occur via the PIK3CA and AKT pathway. Given the inverse relationship between MED7 and EGFR, MED7 may reduce EGFR-mediated ligand-independent ER activation. Its inverse relationship to PIK3CA is also another interesting link to investigate further within the milieu of intersecting ER regulating pathways as other studies indicate that PIK3CA mutations are strongly associated with ER-positive tumours with better prognostic characteristics (Dumont *et al*, 2012). ER+ BCs undergo

extensive chromatin remodeling and histone modifications for hormone-responsive gene expression. For instance, over expression of H4K12ac was associated with ER+ cells and these levels were further increased by estrogen treatment (Nagarajan *et al*, 2015).

High levels of *MED7* mRNA or protein was associated with a better prognosis in BC. On both univariate and multivariate analysis, *MED7* expression was significantly associated with an improved long-term prognosis. In terms of the poor prognostic indicator, LVI (Lee *et al*, 1990), *MED7* was negatively related with LVI, implying its protective role in BC. Given the strong association between *MED7* and ER positive low grade luminal BCs, it is more likely that this correlation is a passenger effect rather than a driver event. Also, the overall correlation with good prognosis in the whole cohort seems to stem chiefly from *MED7*'s strong prognostic correlations in ER+ luminal A tumours. Though the prognostic effect of *MED7* is not observed in ER-ve tumours, the prognostic value in ER positive BCs is potentially useful. Some ER positive tumours are known to recur in the long run; as *MED7* is of prognostic significance over a long time span, this may help discriminate between good vs poorly performing ER positive tumours. In this study, the negative correlations with N-cadherin, CK5/6, and CK17 indicates that *MED7* expression is not associated with aggressive BCs. N-cadherin gain is an epithelial mesenchymal transition (EMT) associated phenomenon contributing to BC aggressiveness (Hulit *et al*, 2007) and tumour invasion and *MED7* may be protective against BC cells acquiring an EMT prone phenotype. Cytokeratins were strongly associated with the high histological grade (III), ER-, PgR- status and worse patient outcome (Alshareeda *et al*, 2013) and their negative association with *MED7* further strengthens its role in non-basal type BC.

There is an increasing focus towards the role of molecular approaches in the classification of BC as well as the use of epigenetics to devise new prognostic markers and predictive tools (Szyf, 2012). Recent research has indicated that epigenetic alterations such as DNA methylation and histone modifications play a role in the development of various cancers including BC, as these changes can

affect multiple gene networks and are able to influence many cellular processes related to tumorigenesis, (Dumitrescu, 2012). Histone modification methods, acetylation (Parthun, 2007) and methylation (Xiao *et al*, 2003), in particular are known to impact on gene expression in cancer including BC (Elsheikh *et al*, 2009). MED7 was significantly positively associated with modified histone marks, H3K4Me2 and H4K12ac (methylation of lysine) and H4R3Me2 (methylation of arginine. High H4R3Me2, with which MED7 was significantly associated, was associated with good prognosis and a longer disease free survival and with luminal-type tumours and hormone receptor expression (Elsheikh *et al*, 2009). Positive vascular invasion, associated with lower MED7 levels, is also known to be associated with low levels of other histone marks like H4K16ac (Elsheikh *et al*, 2009); It may be conjectured that MED7 may be involved in some part of the histone modification process or is recruited to genes which have been modified by histones to confer a better overall prognosis.

This study revealed and confirmed that, MED7 was associated with good prognostic characteristics and better long-term survival outcome in BC. Morphologically, it is significantly associated with invasive lobular cancers. Over-expression of MED7 particularly appears to play a significant role in ER+ Luminal subtype of breast cancer and given its association with multiple ER related markers, further functional assessment is necessary to reveal the specific role played by this mediator protein in these ER positive tumours. The current study suggests a multi-functional role of MED7 in invasive BC biology, and validates the utility of multi-platform approaches (global expression profiling, complemented by immunohistochemistry) in prognostic biomarker discovery.

Acknowledgments

AM acknowledges the NIHR, the Academy of Medical Sciences and the Pathological Society of GB and Ireland for support.

The authors thank the Nottingham Health Science Biobank, Nottingham City Hospital NHS trust, and the Breast Cancer Now Tissue Bank for the provision of tissues used in this study.

Authors' contributions

CJ participated in experimentation, analysis, interpretation, and manuscript drafting. OM conducted the immunohistochemical studies and participated in the analysis and interpretation. MC and SS helped in data management and interpretation; RR carried out the molecular genetics analysis; EP helped with pathology review and manuscript drafting; CCN & MDR helped with the TMA sections; MA helped in immune-histochemical analysis and interpretation; IOE, AG and EAR participated in interpretation and manuscript drafting. AM conceived and supervised the study, participated in its design, interpretation and analysis, including drafting. All authors read and approved the final manuscript.

Titles and Legends to Figures

Figure 1. **A)** MED7 mRNA and PAM50 subtypes, **B)** MED7 mRNA vs BCSS in the whole cohort, **C)** Western Blotting analysis using MED7 Rabbit monoclonal antibody [EPR15410]. Intensity levels of staining are shown: **D)** low, **E)** strong and **F)** negative expression (x200 magnification).

Figure 2. Kaplan-Meier survival plots for MED7 nuclear expression: **A)** MED7 vs BCSS in all cases; **B)** MED7 vs BCSS in Luminal A; **C)** MED7 vs BCSS in Luminal B; **D)** MED7 vs BCSS in TPN; **E)** MED7 vs BCSS in HER2+; **F)** MED7 vs MFI in all cases and **G)** Targeted prognostic analyses for MED7 via the BC gene miner in ER+ node negative patients using Breast Cancer Gene-Expression Miner v4.0.

Supplementary Figure 1: MED7 gene expression and breast cancer patient outcome in different BC sub groups. **A)** MED7 mRNA vs BCSS in Luminal A tumours, **B)** MED7 mRNA vs BCSS in Luminal B tumours, **C)** MED7 mRNA vs BCSS in Triple Negative tumours and **D)** MED7 mRNA vs BCSS in HER2+ tumours.

Supplementary Figure 2: MED7 gene expression and its association, using Breast Cancer Gene-Expression Miner v4.0. **A)** ER status, **B)** PR status and **C)** HER2 status.

REFERENCES

Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JFR, Macmillan D, Blamey RW, Ellis IO (2005) High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *International journal of cancer* **116**(3): 340-350

Abdel-Fatah TM, Powe DG, Hodi Z, Reis-Filho JS, Lee AH, Ellis IO (2008) Morphologic and molecular evolutionary pathways of low nuclear grade invasive breast cancers and their putative precursor lesions: further evidence to support the concept of low nuclear grade breast neoplasia family. *The American journal of surgical pathology* **32**(4): 513-23

Al-Dhaheri M, Wu JC, Skliris GP, Li J, Higashimoto K, Wang YD, White KP, Lambert P, Zhu YR, Murphy L, Xu W (2011) CARM1 Is an Important Determinant of ER alpha-Dependent Breast Cancer Cell Differentiation and Proliferation in Breast Cancer Cells. *Cancer Res* **71**(6): 2118-2128

Aleskandarany MA, Abduljabbar R, Ashankyty I, Elmouna A, Jerjees D, Ali S, Buluwela L, Diez-Rodriguez M, Caldas C, Green AR, Ellis IO, Rakha EA (2016a) Prognostic significance of androgen receptor expression in invasive breast cancer: transcriptomic and protein expression analysis. *Breast cancer research and treatment* **159**(2): 215-27

Aleskandarany MA, Agarwal D, Negm OH, Ball G, Elmouna A, Ashankyty I, Nuglozeh E, Fazaludeen MF, Diez-Rodriguez M, Nolan CC, Tighe PJ, Green AR, Ellis IO, Rakha EA (2016b) The prognostic significance of STAT3 in invasive breast cancer: analysis of protein and mRNA expressions in large cohorts. *Breast cancer research and treatment* **156**(1): 9-20

Aleskandarany MA, Green AR, Benhasouna AA, Barros FF, Neal K, Reis JS, Ellis IO, Rakha EA (2012) Prognostic value of proliferation assay in the luminal, HER2-positive, and triple-negative biologic classes of breast cancer. *Breast Cancer Res* **14**(1)

Aleskandarany MA, Green AR, Rakha EA, Mohammed RA, Elsheikh SE, Powe DG, Paish EC, Macmillan RD, Chan S, Ahmed SI, Ellis IO (2010a) Growth fraction as a predictor of response to chemotherapy in node-negative breast cancer. *International journal of cancer* **126**(7): 1761-1769

Aleskandarany MA, Rakha EA, Ahmed MA, Powe DG, Ellis IO, Green AR (2011) Clinicopathologic and molecular significance of phospho-Akt expression in early invasive breast cancer. *Breast cancer research and treatment* **127**(2): 407-16

Aleskandarany MA, Rakha EA, Ahmed MA, Powe DG, Paish EC, Macmillan RD, Ellis IO, Green AR (2010b) PIK3CA expression in invasive breast cancer: a biomarker of poor prognosis. *Breast cancer research and treatment* **122**(1): 45-53

Alshareeda AT, Soria D, Garibaldi JM, Rakha E, Nolan C, Ellis IO, Green AR (2013) Characteristics of basal cytokeratin expression in breast cancer. *Breast cancer research and treatment* **139**(1): 23-37

Barros FFT, Abdel-Fatah TMA, Moseley P, Nolan CC, Durham AC, Rakha EA, Chan S, Ellis IO, Green AR (2014) Characterisation of HER heterodimers in breast cancer using in situ proximity ligation assay. *Breast cancer research and treatment* **144**(2): 273-285

Ciriello G, Sinha R, Hoadley KA, Jacobsen AS, Reva B, Perou CM, Sander C, Schultz N (2013) The molecular diversity of Luminal A breast tumors. *Breast cancer research and treatment* **141**(3): 409-20

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Group M, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**(7403): 346-52

Dawson SJ, Rueda OM, Aparicio S, Caldas C (2013) A new genome-driven integrated classification of breast cancer and its implications. *The EMBO journal* **32**(5): 617-28

Dumitrescu RG (2012) DNA methylation and histone modifications in breast cancer. *Ras Signaling: Methods and Protocols* **863**: 35-45

Dumont AG, Dumont SN, Trent JC (2012) The favorable impact of PIK3CA mutations on survival: an analysis of 2587 patients with breast cancer. *Chin J Cancer* **31**(7): 327-334

Elsheikh S, Green AR, Aleskandarany MA, Grainge M, Paish CE, Lambros MBK, Reis JS, Ellis IO (2008) CCND1 amplification and cyclin D1 expression in breast cancer and their relation with proteomic subgroups and patient outcome. *Breast cancer research and treatment* **109**(2): 325-335

Elsheikh SE, Green AR, Rakha EA, Powe DG, Ahmed RA, Collins HM, Soria D, Garibaldi JM, Paish CE, Ammar AA, Grainge MJ, Ball GR, Abdelghany MK, Martinez-Pomares L, Heery DM, Ellis IO (2009) Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res* **69**(9): 3802-9

ElTanani MKK, Green CD (1997) Two separate mechanisms for ligand-independent activation of the estrogen receptor. *Mol Endocrinol* **11**(7): 928-937

Fidalgo F, Rodrigues TC, Pinilla M, Silva AG, Maciel Mdo S, Rosenberg C, de Andrade VP, Carraro DM, Krepschi AC (2015) Lymphovascular invasion and histologic grade are associated with specific genomic profiles in invasive carcinomas of the breast. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **36**(3): 1835-48

Finlin BS, Gau CL, Murphy GA, Shao HP, Kimel T, Seitz RS, Chiu YF, Botstein D, Brown PO, Tamanoi F, Andres DA, Perou CM (2001) RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. *J Biol Chem* **276**(45): 42259-42267

Gustafsson CM, Samuelsson T (2001) Mediator--a universal complex in transcriptional regulation. *Molecular microbiology* **41**(1): 1-8

Habashy HO, Powe DG, Glaab E, Ball G, Spiteri I, Krasnogor N, Garibaldi JM, Rakha EA, Green AR, Caldas C, Ellis IO (2011) RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: a marker of ER-positive luminal-like subtype. *Breast cancer research and treatment* **128**(2): 315-326

Habashy HO, Rakha EA, Ellis IO, Powe DG (2013) The oestrogen receptor coactivator CARM1 has an oncogenic effect and is associated with poor prognosis in breast cancer. *Breast cancer research and treatment* **140**(2): 307-316

Haltas H, Bayrak R, Yenidunya S, Kosehan D, Sen M, Akin K (2012) Invasive lobular carcinoma with extracellular mucin as a distinct variant of lobular carcinoma: a case report. *Diagnostic pathology* **7**: 91

Hasegawa N, Sumitomo A, Fujita A, Aritome N, Mizuta S, Matsui K, Ishino R, Inoue K, Urahama N, Nose J, Mukohara T, Kamoshida S, Roeder RG, Ito M (2012) Mediator subunits MED1 and MED24 cooperatively contribute to pubertal mammary gland development and growth of breast carcinoma cells. *Molecular and cellular biology* **32**(8): 1483-95

Hulit J, Suyama K, Chung S, Keren R, Agiostratidou G, Shan W, Dong X, Williams TM, Lisanti MP, Knudsen K, Hazan RB (2007) N-cadherin signaling potentiates mammary tumor metastasis via enhanced extracellular signal-regulated kinase activation. *Cancer Res* **67**(7): 3106-16

Hur K, Lee HJ, Woo JH, Kim JH, Yang HK (2010) Gene expression profiling of human gastrointestinal stromal tumors according to its malignant potential. *Digestive diseases and sciences* **55**(9): 2561-7

Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43**(1): 27-U42

Jezequel P, Campone M, Gouraud W, Guerin-Charbonnel C, Leux C, Ricolleau G, Campion L (2012) bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast cancer research and treatment* **131**(3): 765-75

Karaman S, Detmar M (2014) Mechanisms of lymphatic metastasis. *The Journal of clinical investigation* **124**(3): 922-8

Koschubs T, Seizl M, Lariviere L, Kurth F, Baumli S, Martin DE, Cramer P (2009) Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *The EMBO journal* **28**(1): 69-80

Kwon JY, Park JM, Gim BS, Han SJ, Lee J, Kim YJ (1999) Caenorhabditis elegans mediator complexes are required for developmental-specific transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America* **96**(26): 14990-5

Laganier J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguere V (2005) From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proceedings of the National Academy of Sciences of the United States of America* **102**(33): 11651-6

Lee AKC, Delellis RA, Silverman ML, Heatley GJ, Wolfe HJ (1990) Prognostic-Significance of Peritumoral Lymphatic and Blood-Vessel Invasion in Node-Negative Carcinoma of the Breast. *J Clin Oncol* **8**(9): 1457-1465

Lien HC, Huang CS, Yang YW, Jeng YM (2016) MED12 exon 2 mutation as a highly sensitive and specific marker in distinguishing phyllodes tumours from other spindle neoplasms of the breast. *Apmis* **124**(5): 356-364

Masuda H, Zhang D, Bartholomeusz C, Doihara H, Hortobagyi GN, Ueno NT (2012) Role of epidermal growth factor receptor in breast cancer. *Breast cancer research and treatment* **136**(2): 331-45

McCarty KS, Jr., McCarty KS, Sr. (1984) Histochemical approaches to steroid receptor analyses. *Seminars in diagnostic pathology* **1**(4): 297-308

Mohammed RAA, Martin SG, Gill MS, Green AR, Paish EC, Ellis IO (2007) Improved methods of detection of lymphovascular invasion demonstrate that it is the predominant method of vascular invasion in breast cancer and has important clinical consequences. *Am J Surg Pathol* **31**(12): 1825-1833

Nagarajan S, Benito E, Fischer A, Johnsen SA (2015) H4K12ac is regulated by estrogen receptor-alpha and is associated with BRD4 function and inducible transcription. *Oncotarget* **6**(9): 7305-7317

Parthun MR (2007) Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**(37): 5319-28

Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO (2007) Prognostic markers in triple-negative breast cancer. *Cancer* **109**(1): 25-32

Rakha EA, El-Sayed ME, Menon S, Green AR, Lee AH, Ellis IO (2008) Histologic grading is an independent prognostic factor in invasive lobular carcinoma of the breast. *Breast cancer research and treatment* **111**(1): 121-7

Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA, Evans AJ, Blamey R, Reis JS, Foulkes WD, Ellis IO (2009) Triple-Negative Breast Cancer: Distinguishing between Basal and Nonbasal Subtypes. *Clin Cancer Res* **15**(7): 2302-2310

Rakha EA, Martin S, Lee AH, Morgan D, Pharoah PD, Hodi Z, Macmillan D, Ellis IO (2012) The prognostic significance of lymphovascular invasion in invasive breast carcinoma. *Cancer* **118**(15): 3670-80

Ryu SJ, Zhou S, Ladurner AG, Tjian R (1999) The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* **397**(6718): 446-450

Schiano C, Casamassimi A, Rienzo M, de Nigris F, Sommese L, Napoli C (2014) Involvement of Mediator complex in malignancy. *Biochimica et biophysica acta* **1845**(1): 66-83

Silwal-Pandit L, Vollan HKM, Chin SF, Rueda OM, McKinney S, Osako T, Quigley DA, Kristensen VN, Aparicio S, Borresen-Dale AL, Caldas C, Langerod A (2014) TP53 Mutation Spectrum in Breast Cancer Is Subtype Specific and Has Distinct Prognostic Relevance. *Clin Cancer Res* **20**(13): 3569-3580

Stingl J, Caldas C (2007) Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nature reviews Cancer* **7**(10): 791-9

Szyf M (2012) DNA methylation signatures for breast cancer classification and prognosis. *Genome Med* **4**

Tebbji F, Chen Y, Richard Albert J, Gunsalus KT, Kumamoto CA, Nantel A, Sellam A, Whiteway M (2014) A functional portrait of Med7 and the mediator complex in *Candida albicans*. *PLoS genetics* **10**(11): e1004770

Voduc D, Cheang M, Nielsen T (2008) GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. *Cancer Epidem Biomar* **17**(2): 365-373

Woychik NA, Hampsey M (2002) The RNA polymerase II machinery: Structure illuminates function. *Cell* **108**(4): 453-463

Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell S, Taylor IA, Blackburn GM, Gamblin SJ (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**(6923): 652-6

Yu J, Dabbs DJ, Shuai Y, Niemeier LA, Bhargava R (2011) Classical-type invasive lobular carcinoma with HER2 overexpression: clinical, histologic, and hormone receptor characteristics. *American journal of clinical pathology* **136**(1): 88-97

Zhang L, Cui J, Leonard M, Nephew K, Li Y, Zhang X (2013) Silencing MED1 sensitizes breast cancer cells to pure anti-estrogen fulvestrant in vitro and in vivo. *PLoS one* **8**(7): e70641

Table 1: Associations between MED7 mRNA expression and clinicopathological variables in the METABRIC cohort

Clinicopathological Criteria	MED7 mRNA expression		χ^2 (p Value)
	Low (%)	High (%)	
Age at Diagnosis			
≤50	251 (59.3)	172 (40.7)	17.35(<0.00001)
>50	723 (47.9)	787 (52.1)	
Tumour Size (cm)			
≤2.0	423 (49.4)	434 (50.6)	0.36 (0.554)
>2.0	558 (50.7)	542 (49.3)	
Histological Grade			
1	62 (36.5)	108 (63.5)	21.68 (<0.0001)
2	374 (48.6)	396 (51.4)	
3	521(54.8)	429 (45.2)	
Tumour Type			
Ductal	866 (51.0)	833 (49)	19.39 (0.001)
Lobular*	62 (42.2)	85 (57.8)	
Medullary-like	26 (81.2)	6 (18.8)	
Special type *	19 (38.8)	30 (61.2)	
Miscellaneous	12 (57.1)	9 (42.9)	
NPI			
Good Prognostic Group	313 (46.0)	367(54.0)	7.95 (0.019)
Moderate Prognostic Group	580 (52.8)	519(47.2)	
Poor Prognostic Group	104 (52.3)	95(47.7)	
PAM50 subtype			
Luminal A	297(41.4)	421(58.6)	165.80 (<0.00001)
Luminal B	179 (36.7)	309 (63.3)	
Basal	242 (73.8)	86 (26.2)	
Her2	164 (68.3)	76 (31.7)	
Normal like	112 (56.6)	86 (43.4)	
IntClustMemb			
IntClustMemb 1	62 (44.6)	77 (55.4)	120.96 (<0.00001)
IntClustMemb 2	32 (44.4)	40 (55.6)	
IntClustMemb 3	107 (36.9)	183 (63.1)	
IntClustMemb 4	183 (53.4)	160 (46.6)	
IntClustMemb 5	117 (61.6)	73 (38.4)	
IntClustMemb 6	39 (45.9)	46 (54.1)	

IntClustMemb 7	106 (55.8)	84 (44.2)	
IntClustMemb 8	105 (35.1)	194 (64.9)	
IntClustMemb 9	77 (53.1)	68 (46.9)	
IntClustMemb 10	169 (75.1)	56 (24.9)	
ER			
Negative	312 (71.2)	126 (28.8)	97.43 (< 0.00001)
Positive	665 (44.4)	832 (55.6)	
PR			
Negative	566 (60.3)	373 (39.7)	69.69 (< 0.00001)
Positive	431(41.5)	608 (58.5)	
HER2			
Negative	844 (48.8)	887 (51.2)	15.03 (0.0001)
Positive	153 (61.9)	94 (38.1)	

Significant p values are highlighted in bold.

Table 3: Associations of nuclear MED7 IHC expression and other tissue biomarkers within the Nottingham BC series.

BC Biomarker	MED7 Nuclear Staining		χ^2 (p value)
	Low (%)	High (%)	
ER			
Negative	208 (62.8)	123 (37.2)	28.120 (<0.0001)
Positive	429(45.9)	506 (54.1)	
PgR			
Negative	305 (58.7)	215 (41.3)	22.014 (<0.0001)
Positive	316 (45.1)	385 (54.9)	
HER2			
Negative	517 (49.3)	532 (50.7)	3.22 (0.085)
Positive	98 (56.6)	75 (43.4)	
CK5/6			
Negative	385 (50.4)	379 (49.6)	12.895 (<0.001)
Positive	99 (66.4)	50 (33.6)	
CK17			
Negative	376 (51.8)	350 (48.2)	9.012 (0.003)
Positive	72 (67.3)	35 (32.7)	
STAT3			
Negative	356 (57.7)	261 (42.3)	44.140 (<0.0001)
Positive	77(32.4)	161 (67.6)	
GATA3			
Negative	343(62.4)	207 (37.6)	41.045 (<0.0001)
Positive	59 (34.5)	112 (65.5)	
EGFR			
Negative	468(48.4)	498(51.6)	7.052 (0.009)
Positive	149(57.8)	109(42.2)	
PIK3CA			
Negative/Low	99(43.2)	130(56.8)	13.287 (0.001)
Medium	115(47.7)	126(52.3)	
High	270(57.0)	204(43.0)	
N Cadherin			
Negative	106 (37.7)	175 (62.3)	31.728 (<0.0001)
Positive	355 (58.0)	257 (42.0)	
Ki67			
Low	179 (44.8)	221 (55.2)	10.305 (0.002)
High	328 (55.1)	267 (44.9)	
H3K4Me2			
Negative	82 (54.3)	69 (45.7)	4.591(0.041)
Positive	115 (43.4)	150(56.6)	
H4K12ac			
Negative	89 (56.7)	68 (43.3)	8.654 (0.004)

Positive	107 (41.8)	149 (58.2)	
H4R3Me2			
Negative	113 (55.1)	92 (44.9)	4.233 (0.048)
Positive	115(45.5)	138(54.5)	
RERG			
Negative	326(53.4)	284(46.6)	5.893 (0.015)
Positive	89(43.6)	115(56.4)	
CARM1			
Negative	127(59.1)	88(40.9)	6.057 (0.015)
Positive	85(46.7)	97(53.3)	
FOXA1			
Negative	278(63.5)	160(36.5)	41.623 (< 0.0001)
Positive	154(40.8)	223(59.2)	

Significant p values are highlighted in bold.

Table 2: Relationship between nuclear MED7 protein (IHC) and clinicopathological parameters of the Nottingham BC series.

Clinicopathological Criteria	MED7 Nuclear Staining		χ^2 (p Value)
	Low (%)	High (%)	
Age at Diagnosis			
≤50	220 (48.6)	233 (51.4)	0.781(0.381)
>50	423(51.1)	404 (48.9)	
Tumour Size (cm)			
≤2.0	274 (44.1)	348 (55.9)	19.09 (<0.0001)
>2.0	366 (56.3)	284 (43.7)	
Histological Grade			
1	82 (39.8)	124 (60.2)	49.341(<0.0001)
2	174 (40.8)	252(59.2)	
3	383 (60.2)	253 (39.8)	
Tubule formation			
1	26 (39.4)	40 (60.6)	15.589 (0.0004)
2	185 (43.8)	237 (56.2)	
3	406 (54.6)	338 (45.4)	
Nuclear Pleomorphism			
1	8 (26.7)	22 (73.3)	29.894 (<0.0001)
2	195 (41.9)	270 (58.1)	
3	413 (56.2)	322 (43.8)	
Mitotic score			
1	164 (39.8)	248 (60.2)	43.005 (<0.0001)
2	110 (44.4)	138 (55.6)	
3	343 (60.0)	229 (40.0)	
Tumour Type			
Ductal	551 (51.6)	516 (48.4)	20.575 (0.001)
Lobular*	41 (35.3)	75 (64.7)	
Medullary-like	22 (75.9)	7 (24.1)	
Special type	22 (42.3)	30 (57.7)	
**Lymph Node Stage			
I	398 (51.1)	381 (48.9)	1.853 (0.603)
II	191 (49.6)	194 (50.4)	
III	50 (48.1)	54 (51.9)	
NPI			
Good Prognostic Group	157(40.6)	230 (59.4)	23.546 (<0.0001)
Moderate Prognostic Group	358 (53.1)	316 (46.9)	
Poor Prognostic Group	125 (59.2)	86 (40.8)	
IHC-validated LVI			

Negative	319 (47.5)	352 (52.5)	4.10 (0.04)
Positive	196 (54.1)	166 (45.9)	

*Tumour type p value reflects association between MED7 and lobular tumours

**Lymph node stages 1, 2 and 3 refer to the lymph node staging score incorporated in the Nottingham Prognostic Index, routinely used for breast cancer prognostication: 1: refers to no lymph nodes being involved; 2 refers to 1-3 lymph nodes positive and 3: >3 lymph nodes positive.

Significant p values are highlighted in bold.

Table 4: Univariate and multivariate analysis: Effects of nuclear MED7 expression, LN Stage, Grade and LVI.

Variable	Univariate			Multivariate		
	HR	95%CI	p value	HR	95%CI	p value
Whole cohort						
Stage	2.1	1.9-2.4	<0.001	1.8	1.5 – 2.1	<0.001
Grade	2.3	2.0-2.6	<0.001	1.9	1.6 – 2.3	<0.001
LVI	2.1	1.8-2.6	<0.001	1.5	1.2 – 1.9	0.001
MED7	0.7	0.5-0.8	<0.001	0.7	0.6-0.9	0.007
Luminal A						
Stage	2.3	1.6-3.4	<0.0001	1.9	1.1-3.3	0.025
Grade	2.4	1.6-3.6	<0.0001	1.4	0.8-2.5	0.204
LVI	1.0	0.5-2.0	0.937	0.8	0.4-1.9	0.663
MED7	0.4	0.2 – 0.8	0.012	0.5	0.2 – 0.9	0.028
Luminal B						
Stage	1.8	1.5-2.2	<0.001	1.8	1.3-2.5	<0.001
Grade	1.8	1.4-2.3	<0.001	2.9	0.9-8.7	0.015
LVI	2.0	1.5-2.7	<0.001	1.5	0.9-2.5	0.062
MED7	0.7	0.4-0.9	0.015	0.9	0.6-1.4	0.074
TPN						
Stage	1.7	1.3--2.2	<0.001	1.6	1.1-2.3	0.022
Grade	1.8	1.4-2.3	0.964	1.9	0.8-2.0	0.812
LVI	2.2	1.4-3.3	<0.001	1.7	0.9-3.0	0.085
MED7	0.8	0.5-1.3	0.444	0.9	0.5-1.5	0.673
HER2						
Stage	2.4	1.8-3.1	<0.001	1.8	1.2-2.7	0.002
Grade	2.9	0.9-8.7	0.057	1.5	0.9-2.7	0.141
LVI	2.1	1.3-3.3	0.003	2.4	1.3-4.3	0.003
MED7	1.0	0.6-1.6	0.885	1.0	0.6-1.6	0.868

Significant p values are highlighted in bold.

Table Legends:

Table 1: Associations between MED7 mRNA expression and clinicopathological variables in the METABRIC cohort; Significant p values are highlighted in bold.

Table 2: Relationship between nuclear MED7 protein (IHC) and clinicopathological parameters of the Nottingham BC series. *Tumour type p value reflects association between MED7 and lobular tumours ; **Lymph node stages 1, 2 and 3 refer to the lymph node staging score incorporated in the Nottingham Prognostic Index, routinely used for breast cancer prognostication: 1: refers to no lymph nodes being involved; 2 refers to 1-3 lymph nodes positive and 3: >3 lymph nodes positive. Significant p values are highlighted in bold.

Table 3: Associations of nuclear MED7 IHC expression and other tissue biomarkers within the Nottingham BC series. Significant p values are highlighted in bold.

Table 4: Univariate and multivariate analysis: Effects of nuclear MED7 expression, LN Stage, Grade and LVI. Significant p values are highlighted in bold.

Supplementary Table1: Summary of clinico-pathological characteristics for the METABRIC cohort.

Supplementary Table2: Datasets (Study code and reference) included in the online analysis for MED7 on the BC-GenExMiner v4.0 (Breast Cancer Gene-Expression Miner v4.0 online dataset).

Supplementary Table 3: Summary of clinicopathological data for the Nottingham Tenovus Primary Series

Supplementary Table 4: Biomarkers characterised in the Nottingham Tenovus Primary Series with antibodies used, source, dilutions and cut-offs.