Prognostic Significance of Androgen Receptor Expression in Invasive Breast Cancer: Transcriptomic and Protein Expression Analysis

Mohammad A Aleskandarany[1,2*],Rezvan Abduljabbar[3*], Ibraheem Ashankyty[4], Ahmed Elmouna[4], Dena Jerjees[1], Simak Ali[5], Laki Buluwela[5], Maria Diez-Rodriguez[1], Carlos Caldas[6], Andrew R Green[1], Ian O Ellis[1], Emad A Rakha[1,2]

[1]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.
[2] Department of Pathology, Faculty of Medicine, Menoufia University, Egypt
[3] Department of Oncology, Azadi Teaching Hospital, 1014 AM, Duhok, Kurdistan, Iraq
[4]Molecular Diagnostics and Personalised Therapeutics Unit, University of Ha'il, Ha'il, Saudi Arabia,
[5] Department of Surgery and Cancer, Imperial College London, London, W12 0NN, UK
[6] Centre for the Cambridge Experimental Cancer Medicine Centre (ECMC) and Cambridge Breast Cancer Research Unit.

***: First joint authors**

Correspondence

Mohammed Aleskandarany, *MD, PhD,*
Molecular Pathology Research Unit,
Division of Cancer and Stem Cells,
Nottingham City Hospitals,
The University of Nottingham,
Nottingham NG5 1PB
T: +44 (0) 115 8231859
F: +44 (0) 115 9627768
Email: mohammed.aleskandarany@nottingham.ac.uk

Key words: Breast cancer, Androgen receptor, METABRIC, molecular, clinical, outcome.

Abstract

Background: Differential prognostic roles of Androgen Receptor (AR) have been proposed in breast cancer (BC) depending on tumour oestrogen receptor (ER) status. This study aimed to evaluate the prognostic and/or predictive significance of androgen receptor (AR) expression in invasive BC.

Methods: In this study AR expression was studied on a large (n=1141) consecutive series of early-stage (I-III) BC using tissue microarray and immunohistochemistry (IHC). *AR* mRNA expression was assessed in a subset of cases. The prognostic impact of *AR* mRNA expression was externally validated using the online BC gene expression data sets (n=25 data sets, 4,078 patients).

Results: Nuclear AR IHC expression was significantly associated with features of good prognosis including older age, smaller tumour size, lower grade and lobular histology particularly in the ER-positive tumours. AR was associated with ER-related markers GATA3, FOXa1, RERG and BEX1. Negative association was observed with HER2, p53, Ki67, TK1, CD71 and AGTR1. AR Overexpression was associated with longer survival (p <0.001), independent of tumour size, grade, stage (p=0.033, hazard ratio (HR) =0.80 95%CI=0.64-0.98). Similar associations were maintained in ER+ tumours in univariate and multivariate analysis (p<0.01) both in patients with and without adjuvant endocrine or chemotherapy. *AR* mRNA expression showed significant association with tumour grade, molecular subtypes, and longer 10 and 15 years survival in luminal BC. In the external validation cohorts, *AR* gene expression data was associated with improved patients' outcome (p<0.001, HR=0.84, 95% CI 0.79-0.90).

In conclusion: AR is an independent prognostic factor in ER-positive luminal breast cancer but is also expressed in ER-negative tumours. AR could act as a molecular target in patients with ER-positive disease predicting response to adjuvant therapy.

Introduction

Androgen receptor (AR), similar to oestrogen receptor (ER) and progesterone receptor (PR), belongs to the steroid nuclear receptor family (nuclear receptor 3, group C, member 4; NR3C4) [1, 2]. AR is activated when bound by its specific ligands which results in conformational receptor changes and then receptor translocation into the nucleus where it undergoes dimerisation. The dimer binds to its specific Hormone Receptor Elements (HREs) [3]. The AR functions mainly as a DNA-binding transcription factor that regulates gene expression [4]. *In-vitro* cell lines studies revealed that AR potently inhibits ER transactivation and proliferation and promotes apoptosis [5, 6]. However the interaction between AR and ER remains unclear. It has been suggested that signalling through AR replaces oestrogen-dependent signalling and exerts a stimulatory effect through the androgen responsive element, thereby stimulating transcription of steroid responsive genes [7].

In breast cancer (BC), AR is expressed in 50-88% of cases [8-14] with an average of 61% in all BCs and 75% in ER-positive tumours [15]. Previous studies have shown that AR expression has the potential to predict disease progression [15, 16], as well as the likelihood and duration of response to therapy when used with medroxyprogesterone acetate [17]. It was shown that reduced AR expression can predict a four-fold increase in the risk of BC related death in ER-positive BC patients [5]. Studies of ER-negative and triple negative (TN) BC report conflicting results regarding the prognostic significance of AR with some authors indicating a good prognostic value[18] while others reported it is associated with worse outcome [14]. Molecular classification of BC has also reported the identification of a molecular apocrine class that is characterised by loss of ER expression but with expression of AR and activation of androgen signalling. In a meta-analysis, AR was associated with favourable prognosis of BC irrespective of ER expression [15].

Although AR is the main biological driver and therapeutic target in prostate cancer patients, its therapeutic targeting has been found to pose an anti-proliferative action in BC patients [19]. The use of androgens as hormonal therapy in BC has shown results that are generally comparable to Tamoxifen [20, 21]. For instance, fluoxymesterone, an androgen agonist, was studied in advanced BC patients decades ago with reported therapeutic responses of 14-53% [22-24]. Taken together, this study aimed at assessing the prognostic and predictive significance of AR expression (protein and mRNA) in BC patients with emphasis on BC molecular subtypes.

Materials and Methods

This study was based on:

I: A retrospective cohort comprising a well characterised series of early invasive (TNM Stage I-III, excluding T3 and T4 tumours) primary operable BC patients presented to Nottingham City Hospital from 1987-1997. Tumours in these patients were 5 cm in diameter or less at time of presentation [25]. Patients were uniformly treated according to standard protocol; primary surgery, with either mastectomy or wide local excision, followed by radiotherapy. Before 1989, patients did not receive systemic adjuvant therapy. After 1989, the adjuvant treatment stratification of the patients in this cohort was based on prognostic and predictive factors including hormone receptor (ER) status, menopausal state and Nottingham Prognostic Index (NPI). Patients with good prognostic index (NPI <3.4) were considered low risk and did not receive adjuvant systemic therapy. Pre-menopausal women with an NPI >3.4 (high risk) received classical cyclophosphamide, methotrexate and 5-flurouracil (CMF) chemotherapy and patients with ER-positive tumours were offered hormone therapy (HT). Post-menopausal women with an NPI score >3.4 and ER-positivity were offered HT whilst ER-negative patients were offered classical CMF chemotherapy [26]

Patients' clinical and pathological data including age, histological tumour type, primary tumour size, lymph node status, histological status, NPI [27], and vascular invasion were available and prospectively maintained. This cohort has been well investigated using a wide range of biological markers of close relevance to BC biology and outcome. These markers include AR (previously stained in this cohort using anti-AR primary antibody, clone F39.4.1, Biogenex, UK, and the Streptavidin-Biotin "ABC" secondary detection method) [25], ER, PR, HER2, cyokeratins (Ck), Cyclin B1, Ki67, and P53. For further insight into the relationship between ER and AR (i.e. ER-AR interaction), the ER-related proteins including

FOXa1, GATA3, BEX1, PELP1, RERG, and TK1 were also used in this analysis. Survival data includes Breast Cancer Specific Survival (BCSS), in months, from the date the primary surgical treatment to the time of death from breast cancer. Distant metastasis free interval (DMFI) was defined as the time, in months, taken from primary surgical treatment to the first distant recurrence. The duration of follow-up for this study was 306 months and the (mean survival = 150, median was 167 months, with 1136 patients had survival data at the end of follow-up time. Clinicopathological data of this series is presented in supplementary Table 1). A subset of this cohort (n=284) was included in the multicentre study; the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study, for which the human genome has been characterised. For this subset, mRNA was extracted from fresh frozen tumours and hybridised to Illumina HT-12 v3 platform (Bead Arrays) and the data were pre-processed and normalised as previously described [28]. The AR mRNA expression was investigated in this subset. Correlation co-efficient was used to test for association between AR immunohistochemical (IHC) expression and *AR* mRNA expression data, while appropriate statistical tests were used to test for association with clinicopathological variables as well as patients' outcome. This research was approved by Nottingham Research Ethics Committee 2 under the title of ''Development of a molecular genetic classification of breast cancer''. All cases included in this study were from patients who were consented prior to inclusion in the study cohort.

This study adheres to REporting recommendations for tumour MARKer prognostic studies (REMARK) criteria [29]

Immunohistochemistry (IHC)

Validating the antibody specificity:

Prior to IHC, the specificity of the anti-AR primary antibody (Sc-816, Santa Cruz Biotechnology, UK) was validated using Western blotting (WB). WB was performed on

whole cell lysates of MCF-7 and MDA-MB 231 human BC cell lines (obtained from the American Type Culture Collection; Rockville, MD, USA) using 1:100 dilution of the primary antibody dilution, and 1:2,000 of the HRP labelled secondary anti-rabbit antibody, as previously described [30]. This showed a single specific band at the predicted size (110 KDa) of AR protein, confirming the specificity of the antibody (Figure 1A).

Procedure of Immunohistochemistry

IHC was applied to tissue microarrays (TMA) using Novolink™ Max Polymer Detection System from Leica Biosystems (Leica, Newcastle, UK). Heat induced retrieval of antigen epitopes was performed in citrate buffer (pH 6) using microwave for 20 minutes. Anti-AR primary antibody (Sc-816, Santa Cruz Biotechnology, UK) was applied, diluted at 1:40, for 60-minutes incubation. For reaction visualisation, 3-3' Diam-inobenzidine tetrahydrochloride (Novolink DAB substrate buffer plus) was used as a chromogen. Negative (primary antibody replaced by Phosphate-buffered saline (PBS) and positive controls (FFPE tissue section of a known AR positive BC case) were included in the staining run.

Assessment of IHC staining

Slides were scanned into high resolution digital images (0.45µm/pixel) using a NanoZoomer slide scanner (Hamamtsu Photonics, Welwyn Garden City, UK) and uploaded into web server where they could be accessed using a web based interface (Distiller, Leica). TMA cores were scored at 20x magnification using a minimum of 24" high resolution computer screen (1920x1080). The semi-quantitative immunohistochemical scoring (H-score) method was used, which takes the intensity and percentage of stained invasive tumour tissue stained into account [31]. All cases were scored without prior knowledge of the patients' pathologic or outcome data. Although, cytoplasmic staining was observed in some TMA cores, only nuclear staining was considered in this study.

Breast cancer molecular subtypes were defined based on their IHC expression profile into: 1) luminal/hormone receptor (HR)+; ER+ and/or PR+, 2) HER2+ (HER2+ regardless of the expression of other markers), 3) Triple negative basal-like; ER-, PR-, HER2-, and positive for CK5/6, and/or CK14 and/or EGFR and 4) triple negative non-basal BC; all above markers negative), as previously described [32]

II: External validation cohorts: For external validation of *AR* mRNA expression, bc-GenExMiner v3.0 (Breast Cancer Gene-Expression Miner v3.0) online dataset (http://bcgenex.centregauducheau.fr) was used. In this study, the "prognostic module", offering the possibility to evaluate the prognostic impact of candidate genes in breast cancer, was used [33]. Cox model, Kaplan–Meier and forest plots were generated.

Statistical Analysis

The statistical analysis was performed using Statistical Package for Social Sciences SPSS version 21 for Windows (SPSS Inc., Chicago, IL, USA). A *p* value (two-tailed) of less than 0.05 was considered significant. Spearman correlation co-efficient was used to test for correlation between continuous variables. Cut-off values for the different biomarkers included in this study were chosen before statistical analysis. Standard cut-offs were used for established prognostic factors and were the same as for previously published patient series [32]. Determination of the optimal AR H-score cut-offs was obtained using X-tile bio-informatics software [34]. Analysis of categorical variables was performed using the appropriate statistical test. Kaplan-Meier plots were used to visualise the survival distribution of dichotomised AR, with differences in survival estimated using Log-rank tests. Multivariate Cox proportional hazards model was fitted to adjust for confounders and test statistical independence of AR in predicting BCSS and DMFS.

Results

I: AR IHC results:

The number of BC cases informative on the TMA for AR expression in this study was 1141. Figure 1 B, C & D displays representative images of the tumour tissue cores with varying degrees of AR staining.

The distributions of AR H-scores did not follow normal distribution. Optimum cut-off point for AR expression, as defined by X-tile software using BCSS as endpoint, was set at H score = 190. This cut-off point was used to categorise cases into AR negative/low (H-score <190; n=528/1141, 46.3%) and AR positive/high (H-score ≥190; n=613/1141, 53.7%) expression.

Associations of AR IHC expression with clinicopathological features and other markers

Nuclear expression of AR showed positive association with patients' age (p <0.001) and pathological features of good prognosis including smaller tumour size (p=0.001), lower histological grade, more tubule formation, less pleomorphism and low mitotic counts, lobular histological type and special tumour types of excellent prognosis, and low NPI score (p<0.001). Negative association was observed with tumours of medullary-like histology. Importantly, AR expression was negatively associated with mitosis in the whole series, irrespective of ER status. Within the whole series, AR was positively associated with luminal enriched proteins FOXa1, GATA3, BEX1, PELP1, and RERG, while negative association was observed with the HER2+ status (p=0.003), P53 status, Ki67LI, and CD71 (p<0.001).

AR was associated with BC molecular intrinsic subtypes as defined by IHC surrogate markers [32] (p < 0.001, and p= 0.034, for the whole series and ER+ tumours, respectively; Table 2). Highest AR expression was observed in luminal tumours (499/778, 64.1%), while it was less expressed in HER2+ and TN Basal-like, while the TN-non-Basal BC showed the least AR expression (59/142, 41.5%, 33/142, 23.2% and 5/48, 10.4%, respectively).

Moreover, within ER+/luminal tumours, AR was positively associated with ER-related and luminal-enriched genes including, AGTR1 (p= 0.014), PELP1 (p= 0.026), CARM1 (p= 0.008), CD71 (p < 0.016), FOXa1, (p <0.001), GATA3 (p <0.001), BEX1 (p <0.001), and RERG (p= 0.002). Only TK1 was significantly associated (p = 0.011) with AR in the ER- tumours (Table 2).

Associations of AR with patients' outcome

High expression of AR was associated with longer BCSS (Log rank (LR) = 17.88, p <0.001; Figure 2A). In ER-positive tumours, higher AR levels were predictive of longer BCSS (LR=14.58; p <0.001; Figure 2B). Cox proportional multivariate analysis showed that higher AR level of expression was an independent indicator of better outcome irrespective of tumour size, grade and nodal stage (p = 0.033, hazard ratio (HR) = 0.80 95% CI = 0.64-0.98; Table 3). Multivariate analysis showed that AR is an independent prognostic marker (p=0.007, HR = 0.71 95% CI = 0.56-0.91; Table 3). However, inclusion of Ki67LI into the cox proportional hazard model rendered AR expression insignificant (p=0.255, and p=0.091, for the whole series and luminal type, respectively).

Regarding distant metastasis, there was an association between low levels of AR expression and probability of development of DM (LR = 23.32; p < 0.001), Figure 2C. This relationship was maintained using a Cox regression model which showed that AR was an independent predictor of longer DMFI (HR=0.80; p=0.034; 95% CI = 0.65-0.98). A similar association was also observed in the luminal subgroup (Figure 2D). Inclusion of Ki67LI into the cox proportional hazard model rendered AR expression statistically insignificant (p = 0.690, and p = 0.550, for the whole series and luminal type, respectively).

The favourable prognostic significance of AR expression was evident in ER-positive patients regardless of their hormonal therapy status. Patients who received, as well as those who did

not receive hormonal therapy (HT), showed significant survival advantage in the AR-positive group, Figures 3A and 3B, respectively. Considering adjuvant chemotherapy, the prognostic advantage of positive AR expression was maintained in patients with ER-positive disease who either received or did not receive adjuvant chemotherapy (LR= 4.62, p = 0.032, LR= 8.06, p = 0.005, respectively, Figures 3C and 3D). Using X-tile, AR was able to stratify BC patients into three prognostic groups; those with AR expression level less than 200 H-score had shorter 15-year BCSS compared to both intermediate-risk group (200-260) and least-risk group (>260) (LR = 18.78; p <0.001).

When the analysis was restricted to the TNBC and HER2+ phenotypes, nuclear AR expression was neither associated with BCSS nor with DMFI even when using different cut-off point for dichotomisation of AR expression (data not shown).

II: *AR* mRNA expression

AR mRNA expression was assessed in a subset of cases that were included in the METABRIC study [35] (n=284 cases). This showed a significant positive correlation between AR protein IHC expression and *AR* mRNA expression (r=0.424, p<0.001). *AR* mRNA showed significantly higher expression in low grade tumours (One-way ANOVA (F), =5.046, p = 0.007) with low proliferative activity as assessed by mitotic scores (F= 8.056, p < 0.001) and Ki67LI (r=0.403, p < 0.001).

Significant differences were observed between BC molecular subtypes (F= 29.361, p < 0.001), with highest expression observed in the luminal/ER+ subtype. Within the luminal/ER+ tumours using the median cut-off, high *AR* mRNA expression showed longer BCSS as compared to low expression, at 10 and 15 of follow-up with a trend towards significant association at 20 years of follow-up (data not shown).

B: External validation cohorts

Using bc-GenExMiner v3.1 (Breast Cancer Gene-Expression Miner v3.1) online dataset as external validation cohorts, the prognostic impact of *AR* mRNA expression was investigated. As shown in the Forest plot (Supplementary Figure 1), 25 datasets were investigated. In 5/25 studies, high *AR* mRNA expression was significantly associated with improved survival, in one dataset (n=155 patients) it was significantly associated with shorter survival, while the rest of data sets did not show significant association with outcome (Supplementary Table 2). When data of all datasets was pooled together (n=4,078), high AR mRNA expression was significantly associated with better metastatic recurrence (MR) free survival ($p<0.001$, HR=0.78, 95%CI 0.69-0.88; Supplementary Figure 1). This association was maintained in the ER-positive datasets ($p= 0.0017$, HR = 0.78, 95%CI 0.67 - 0.91). However, associations of *AR* with improved outcome deemed statistically insignificant when the analysis was adjusted for proliferation both in the whole and ER-positive datasets (n=2,528 patients, $p= 0.8824$ and n= 1,907, $p = 0.6654$, respectively).

Discussion:

Despite being an important member of the steroid receptor superfamily, its frequent expression in BC and its relation to ER, the biological and clinical significance of AR remains under investigation. Although multiple authorities have reported a good prognostic impact of AR expression in BC [36, 37], others have described that patients with AR over-expressing ER-negative BC have shorter survival [38]. Molecular apocrine class, which is characterised by AR expression and pathway activation typically lacks ER expression and is associated with poor prognosis. Moreover, the magnitude of AR expression impact on outcome has been variably reported in different BC molecular subtypes and whether it is dependent on ER or proliferation remains to be defined. In addition, the correlation between *AR* mRNA and AR protein expression and the clinical significance of each remains to be characterised.

In this study, there was high prevalence of AR expression, more than ER and PgR expression, with those cases showing complete absence of AR forming a minority of cases. This goes in line with previous reports [39, 40]. Interestingly, in studies that have utilised smaller number of patients, it has been shown that AR is frequently expressed in some special types of invasive BC more than others such as apocrine and invasive lobular carcinoma [41-44]. Our study supports these findings and demonstrates that AR expression is significantly higher in invasive lobular carcinoma as well as other special histologic types of excellent prognosis. These associations were significant in the whole series, ER+, and ER- subgroups.

Consistent with previous studies [18, 45, 46], our results showed an association between AR expression features of good prognosis including older patient age, small tumour size, lower tumour grade with lower proliferative activity as well as hormone receptor positivity. The

interaction between AR and ER in down-regulation cancer cell proliferation has led some authorities to speculate that combined selective estrogen/androgen hormone modulators may be an alternative promising modality to the current modalities in hormone receptor positive BC [47, 48]. In this study, luminal BC showed the highest expression of AR compared to HER-positive and TN tumours. Moreover, AR was associated with markers known to be regulated via ER and more expressed in luminal BC such as GATA3, FOXa1, CARM1, RERG, PELP1, AGTR1, CD71, and BEX1. From these results, it appears that AR prognostic value of expression varies significantly depending on the ER status of the tumour. In line with notion, it has been suggested that the dependency of AR on ER status is related to the competitive interaction between AR and ER [5], where in the presence of the latter, AR interacts with oestrogen response elements on ER, blocking downstream oestrogen target genes, leading to inhibition of ER stimulated tumour proliferation. Contrasting this, when ER is lacking/absent, AR interacts with AR elements and promotes tumour cell growth [49]. In line with this is the report of *Elebro* and co-authors who reported on the interaction of AR and ER in determining the impact of AR expression [50].

Regarding the association with patients' outcome, AR overexpression was shown to be significantly associated with improved outcome in the whole series and in ER-positive BC. These associations were independent of other well-established prognostic variables including tumour size, grade and nodal stage. However, this prognostic advantage of AR expression was lost upon inclusion of proliferative fraction, as assessed by Ki67LI, into the multivariable model, indicating dependency of AR on tumour proliferation.

In this study, *AR* mRNA expression showed a positive correlation with AR protein expression. Consistent with AR protein expression, higher *AR* mRNA expression was observed in low grade tumours, and was differentially expressed in BC molecular subtypes,

with highest was observed in the luminal/ER-positive subtype conferring better outcome as compared to low expression both as early as late during follow-up period (at 10 and 15 years). Furthermore, 25 online datasets were investigated for *AR* mRNA expression as external validation cohorts. Significantly longer distant recurrence free survival was observed with high *AR* mRNA expression both in unselected (n=4,078) and ER-positive datasets. Once again, these associations deemed statistically insignificant upon adjustment for proliferation in the whole datasets and in ER+ datasets. Therefore, AR prognostic significance appears to be highest in ER+ low proliferative tumours; results in line with AR IHC in our BC series.

It is noteworthy that the improved outcome of high AR IHC expression was maintained in patients received adjuvant hormonal or chemotherapy, or did not receive adjuvant therapy. In both instances of adjuvant therapy, patients with AR positive tumours performed better than those with AR negative tumours. Therefore, adjunctive therapeutic manipulation of AR, with or without other therapies, could be useful in improving outcome, especially in chemotherapy-intolerant patients. Moreover, not only the status of AR as assessed by IHC which could stratify patients, but the level of AR expression was also useful in distinguishing three distinct prognostic groups in BC. Cases which expressed highest levels of AR survived the longest while those with tumours expressed lowest AR levels survived significantly shorter periods. This finding can have useful implications on treatment selection for specific AR expression subgroups of BC patients.

TNBC are regarded as a heterogeneous subtype of BC with chemotherapy as the mainstay of treatment for both early stage and advanced invasive TNBC. In this study 20% of TNBC showed overexpression of AR at H score = 190 cut-off point. However, 85% of cases of TNBC expressed AR showing 1% or more of staining. Accordingly; AR expression was not only expressed in the ER+/luminal tumours but also in the TNBC. However, it was not

significantly associated with patients' outcome in the TNBC neither in the basal phenotype tumours at cut-off point used for the whole series and the ER+ subgroup nor with other cut-offs. Current studies are investigating novel therapeutic strategies in phase II and phase III clinical trials. The anti-androgens for AR positive TNBC are amongst those molecularly targeted therapies currently being tested [51]. Although findings in our data did not show prognostic significance of AR in TNBC, this does not preclude the potential utility of AR molecular targeted therapies for TNBC tumours. It is worth mentioning that variable results of AR expression in literature, both in unselected and different intrinsic molecular subtypes, could be attributed to many reasons. These include the inherent heterogeneous nature of invasive BC, intratumoural heterogeneity of AR expression, using different antibody clones or secondary detection kits with variable sensitivity, or inter-observer variability in interpreting the IHC staining. In our study, the overall agreement between the used antibody in this study and a previously used clone [32] showed excellent agreement. Moreover, it produced the same associations with clinicopathological, biomarker and patients' outcome (data not shown).

The use of such therapy stands as a potential avenue especially in case of treatment failure/resistance with the currently used chemotherapeutic regimens [52, 53]. Meanwhile, several issues that surround the expression of AR in BC should be addressed, such as its relationship with the HER2-pathway, and most importantly if the receptor expressed in BC is different from ARs expressed in other tissues of the body [54, 55]. Should similarities be present, they will be critical to minimise any potential side effects associated with any future AR-targeted therapy.

In conclusion, AR is abundantly expressed in invasive BC with significant association with favourable prognostic parameters both at protein and transcriptomic levels. It is an independent prognostic factor and can further stratify the patient into distinct prognostic subgroups significantly different in outcome. There is interaction between AR and ER and proliferation which may explain the differential prognostic effect of AR in BC. Understanding these biological interactions may help in utilising AR as a molecular therapeutic target in invasive BC.

Acknowledgment:
MA Aleskandarany and Maria Diez-Rodriguez are funded by the University of Ha'il, KSA.

Conflict of interest: The authors have no conflicts of interest to declare.

Titles and legends to figures

Figure 1: Western Blotting (WB) of AR antibody and IHC expression in invasive BC TMA:

A) WB on whole cell lysates of MCF-7 and MDA-MB 231 human BC cell lines using anti-AR primary antibody (Sc-816, Santa Cruz, UK) 1:100 dilution of the primary antibody dilution, and 1:2,000 of the HRP labelled secondary anti-rabbit antibody. Immunohistochemical expression of AR in invasive BC: B) negative AR; C) weak to moderate AR expression; and D) Strong AR expression.

Figure 2: Kaplan Meier plot for the association of AR nuclear expression in relation to BCSS and metastasis free survival:

A) in the whole series; B) in ER-positive tumours. C) Distant metastasis free survival in patients expressing AR in the whole series, and D) in ER+ tumours.

Figure 3: Kaplan Meier plot for the association of AR nuclear expression in relation to Adjuvant therapy:

A) Patients with no hormonal therapy (HT), B) Patients who received HT. C: Patients received adjuvant chemotherapy, and D) patients who did not receive adjuvant chemotherapy.

Supplementary Figure 1: A) Forest plot of *AR* mRNA expression impact on patients' outcome in the 25 external validation cohorts. **B**) Association of AR mRNA expression in the external validation cohorts with patient outcome (= 4078 patients)

Tables legends:

Table 1: Correlations between AR expression and clinico-pathological features in the whole series and in ER-positive BC

Table 2: Association between AR and the expression of other markers in the whole series and in ER-positive BC

Table 3: Cox proportional hazard analysis for predictors of BCSS within the whole studied series and within ER+ tumours only.

Supplementary Table 1: Clinico-pathological features of the study cohort

Supplementary Table 2: The prognostic impact *AR* mRNA in the external validation cohorts using bc-GenExMiner v4.0 for publicly available online datasets (n=25 datasets, 4078 patients).

References:

1. Wilson CM, McPhaul MJ (1994) A and B forms of the androgen receptor are present in human genital skin fibroblasts. *Proc Natl Acad Sci U S A*, 91(4):1234-1238.
2. Gregory CW, He B, Wilson EM (2001) The putative androgen receptor-A form results from in vitro proteolysis. *J Mol Endocrinol*, 27(3):309-319.
3. Heemers HV, Tindall DJ (2007) Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev*, 28(7):778-808.
4. Mooradian AD, Morley JE, Korenman SG (1987) Biological actions of androgens. *Endocr Rev*, 8(1):1-28.
5. Peters AA, Buchanan G, Ricciardelli C, Bianco-Miotto T, Centenera MM, Harris JM, Jindal S, Segara D, Jia L, Moore NL *et al* (2009) Androgen receptor inhibits estrogen receptor-alpha activity and is prognostic in breast cancer. *Cancer Res*, 69(15):6131-6140.
6. Hackenberg R, Schulz KD (1996) Androgen receptor mediated growth control of breast cancer and endometrial cancer modulated by antiandrogen- and androgen-like steroids. *J Steroid Biochem Mol Biol*, 56(1-6 Spec No):113-117.
7. Robinson JL, Macarthur S, Ross-Innes CS, Tilley WD, Neal DE, Mills IG, Carroll JS (2011) Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. *EMBO J*, 30(15):3019-3027.
8. Birrell SN, Butler LM, Harris JM, Buchanan G, Tilley WD (2007) Disruption of androgen receptor signaling by synthetic progestins may increase risk of developing breast cancer. *FASEB J*, 21(10):2285-2293.
9. Moe RE, Anderson BO (2007) Androgens and androgen receptors: a clinically neglected sector in breast cancer biology. *J Surg Oncol*, 95(6):437-439.
10. Gonzalez LO, Corte MD, Vazquez J, Junquera S, Sanchez R, Alvarez AC, Rodriguez JC, Lamelas ML, Vizoso FJ (2008) Androgen receptor expresion in breast cancer: relationship with clinicopathological characteristics of the tumors, prognosis, and expression of metalloproteases and their inhibitors. *BMC Cancer*, 8:149.
11. Ogawa Y, Hai E, Matsumoto K, Ikeda K, Tokunaga S, Nagahara H, Sakurai K, Inoue T, Nishiguchi Y (2008) Androgen receptor expression in breast cancer: relationship with clinicopathological factors and biomarkers. *Int J Clin Oncol*, 13(5):431-435.
12. Isola JJ (1993) Immunohistochemical demonstration of androgen receptor in breast cancer and its relationship to other prognostic factors. *J Pathol*, 170(1):31-35.
13. Niemeier LA, Dabbs DJ, Beriwal S, Striebel JM, Bhargava R (2010) Androgen receptor in breast cancer: expression in estrogen receptor-positive tumors and in estrogen receptor-negative tumors with apocrine differentiation. *Mod Pathol*, 23(2):205-212.
14. Hu R, Dawood S, Holmes MD, Collins LC, Schnitt SJ, Cole K, Marotti JD, Hankinson SE, Colditz GA, Tamimi RM (2011) Androgen receptor expression and breast cancer survival in postmenopausal women. *Clin Cancer Res*, 17(7):1867-1874.
15. Vera-Badillo FE, Templeton AJ, de Gouveia P, Diaz-Padilla I, Bedard PL, Al-Mubarak M, Seruga B, Tannock IF, Ocana A, Amir E (2014) Androgen

receptor expression and outcomes in early breast cancer: a systematic review and meta-analysis. *J Natl Cancer Inst*, 106(1):djt319.
16. Schippinger W, Regitnig P, Dandachi N, Wernecke KD, Bauernhofer T, Samonigg H, Moinfar F (2006) Evaluation of the prognostic significance of androgen receptor expression in metastatic breast cancer. *Virchows Arch*, 449(1):24-30.
17. Birrell SN, Roder DM, Horsfall DJ, Bentel JM, Tilley WD (1995) Medroxyprogesterone acetate therapy in advanced breast cancer: the predictive value of androgen receptor expression. *J Clin Oncol*, 13(7):1572-1577.
18. 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.
19. Toth-Fejel S, Cheek J, Calhoun K, Muller P, Pommier RF (2004) Estrogen and androgen receptors as comediators of breast cancer cell proliferation: providing a new therapeutic tool. *Arch Surg*, 139(1):50-54.
20. Tormey DC, Lippman ME, Edwards BK, Cassidy JG (1983) Evaluation of tamoxifen doses with and without fluoxymesterone in advanced breast cancer. *Ann Intern Med*, 98(2):139-144.
21. Ingle JN, Twito DI, Schaid DJ, Cullinan SA, Krook JE, Mailliard JA, Tschetter LK, Long HJ, Gerstner JG, Windschitl HE *et al* (1991) Combination hormonal therapy with tamoxifen plus fluoxymesterone versus tamoxifen alone in postmenopausal women with metastatic breast cancer. An updated analysis. *Cancer*, 67(4):886-891.
22. Kennedy BJ (1958) Fluoxymesterone therapy in advanced breast cancer. *N Engl J Med*, 259(14):673-675.
23. Stoll BA (1959) Fluoxymesterone (halotestin) in advanced breast carcinoma. *Med J Aust*, 46(3):70-74.
24. Lowe R, De Lorimier AA, Jr., Gordan GS, Goldman L (1961) Antitumor efficacy of fluoxymesterone. Use in advanced breast cancer. *Arch Intern Med*, 107:241-244.
25. Abd El-Rehim DM, Ball G Fau - Pinder SE, Pinder Se Fau - Rakha E, Rakha E Fau - Paish C, Paish C Fau - Robertson JFR, Robertson Jf Fau - Macmillan D, Macmillan D Fau - Blamey RW, Blamey Rw Fau - Ellis IO, 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. (0020-7136 (Print)).
26. Blamey RW, Ellis IO, Pinder SE, Lee AH, Macmillan RD, Morgan DA, Robertson JF, Mitchell MJ, Ball GR, Haybittle JL *et al* (2007) Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. *Eur J Cancer*, 43(10):1548-1555.
27. Galea MH, Blamey RW, Elston CE, Ellis IO (1992) The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat*, 22(3):207-219.
28. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346-352.

29. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Oncol*, 2(8):416-422.
30. Mruk DD, Cheng CY (2011) Enhanced chemiluminescence (ECL) for routine immunoblotting: An inexpensive alternative to commercially available kits. *Spermatogenesis*, 1(2):121-122.
31. McCarty KS, Jr., McCarty KS, Sr. (1984) Histochemical approaches to steroid receptor analyses. *Seminars in diagnostic pathology*, 1(4):297-308.
32. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA *et al* (2009) Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clin Cancer Res*, 15(7):2302-2310.
33. 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 Res Treat*, 131(3):765-775.
34. Camp RL, Dolled-Filhart M, Rimm DL (2004) X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*, 10(21):7252-7259.
35. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346-352.
36. Alshenawy HA (2012) Prevalence of androgen receptors in invasive breast carcinoma and its relation with estrogen receptor, progesterone receptor and Her2/neu expression. *Journal of the Egyptian National Cancer Institute*, 24(2):77-83.
37. Qu Q, Mao Y, Fei XC, Shen KW (2013) The impact of androgen receptor expression on breast cancer survival: a retrospective study and meta-analysis. *PloS one*, 8(12):e82650.
38. Farmer P, Bonnefoi H, Anderle P, Cameron D, Wirapati P, Becette V, André S, Piccart M, Campone M, Brain E (2009) A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nature medicine*, 15(1):68-74.
39. Kuenen-Boumeester V, Van der Kwast TH, Claassen CC, Look MP, Liem GS, Klijn JG, Henzen-Logmans SC (1996) The clinical significance of androgen receptors in breast cancer and their relation to histological and cell biological parameters. *Eur J Cancer*, 32A(9):1560-1565.
40. Moinfar F, Okcu M, Tsybrovskyy O, Regitnig P, Lax SF, Weybora W, Ratschek M, Tavassoli FA, Denk H (2003) Androgen receptors frequently are expressed in breast carcinomas: potential relevance to new therapeutic strategies. *Cancer*, 98(4):703-711.
41. Gatalica Z (1997) Immunohistochemical analysis of apocrine breast lesions. Consistent over-expression of androgen receptor accompanied by the loss of estrogen and progesterone receptors in apocrine metaplasia and apocrine carcinoma in situ. *Pathol Res Pract*, 193(11-12):753-758.
42. Riva C, Dainese E, Caprara G, Rocca PC, Massarelli G, Tot T, Capella C, Eusebi V (2005) Immunohistochemical study of androgen receptors in breast carcinoma. Evidence of their frequent expression in lobular carcinoma. *Virchows Arch*, 447(4):695-700.

43. Shim HS, Jung WH, Kim H, Park K, Cho NH (2006) Expression of androgen receptors and inhibin/activin alpha and betaA subunits in breast apocrine lesions. *Apmis*, 114(5):352-358.
44. Cho LC, Hsu YH (2008) Expression of androgen, estrogen and progesterone receptors in mucinous carcinoma of the breast. *Kaohsiung J Med Sci*, 24(5):227-232.
45. Agrawal AK, Jelen M, Grzebieniak Z, Zukrowski P, Rudnicki J, Nienartowicz E (2008) Androgen receptors as a prognostic and predictive factor in breast cancer. *Folia Histochem Cytobiol*, 46(3):269-276.
46. Hanley K, Wang J, Bourne P, Yang Q, Gao AC, Lyman G, Tang P (2008) Lack of expression of androgen receptor may play a critical role in transformation from in situ to invasive basal subtype of high-grade ductal carcinoma of the breast. *Hum Pathol*, 39(3):386-392.
47. Tiefenbacher K, Daxenbichler G (2008) The Role of Androgens in Normal and Malignant Breast Tissue. *Breast care (Basel, Switzerland)*, 3(5):325-331.
48. Zhou J, Ng S, Adesanya-Famuiya O, Anderson K, Bondy CA (2000) Testosterone inhibits estrogen-induced mammary epithelial proliferation and suppresses estrogen receptor expression. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 14(12):1725-1730.
49. Hickey TE, Robinson JL, Carroll JS, Tilley WD (2012) Minireview: The androgen receptor in breast tissues: growth inhibitor, tumor suppressor, oncogene? *Molecular endocrinology (Baltimore, Md)*, 26(8):1252-1267.
50. Elebro K, Borgquist S, Simonsson M, Markkula A, Jirstrom K, Ingvar C, Rose C, Jernstrom H (2015) Combined Androgen and Estrogen Receptor Status in Breast Cancer: Treatment Prediction and Prognosis in a Population-Based Prospective Cohort. *Clin Cancer Res*, 21(16):3640-3650.
51. Lehmann BD, Pietenpol JA, Tan AR (2015) Triple-negative breast cancer: molecular subtypes and new targets for therapy. *American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting*, 35:e31-39.
52. Barton VN, D'Amato NC, Gordon MA, Lind HT, Spoelstra NS, Babbs BL, Heinz RE, Elias A, Jedlicka P, Jacobsen BM *et al* (2015) Multiple Molecular Subtypes of Triple-Negative Breast Cancer Critically Rely on Androgen Receptor and Respond to Enzalutamide In Vivo. *Molecular cancer therapeutics*.
53. Gucalp A, Tolaney S, Isakoff SJ, Ingle JN, Liu MC, Carey LA, Blackwell K, Rugo H, Nabell L, Forero A *et al* (2013) Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic Breast Cancer. *Clin Cancer Res*, 19(19):5505-5512.
54. McNamara KM, Moore NL, Hickey TE, Sasano H, Tilley WD (2014) Complexities of androgen receptor signalling in breast cancer. *Endocrine-Related Cancer*, 21(4):T161-T181.
55. Hu DG, Hickey TE, Irvine C, Wijayakumara DD, Lu L, Tilley WD, Selth LA, Mackenzie PI (2014) Identification of androgen receptor splice variant transcripts in breast cancer cell lines and human tissues. *Hormones & cancer*, 5(2):61-71.

Table 1: Associations between AR expression and clinico-pathological features in the whole series and in ER-positive BC

	AR Expression in the whole series			**AR Expression in the ER-Positive tumours**			**AR Expression in the ER-Negative tumours**		
	Negative/low N (%)	Positive/high N (%)	p-value(χ^2)	Negative/low N (%)	Positive/high N (%)	p-value(χ^2)	Negative/low N (%)	Positive/high N (%)	p-value(χ^2)
Age									
<40	60 (63.8)	34 (36.2)		21 (45.7)	25 (54.3)		39 (81.3)	9 (18.8)	
40-49	147 (47.0)	166 (53.0)	**<0.001** (19.93)	82 (35.7)	148 (64.3)	0.118 (5.88)	64 (80.0)	16 (20.0)	0.235 (4.25)
50-59	171 (48.4)	182 (51.6)		108 (41.2)	154 (58.8)		62 (69.7)	27 (30.3)	
≥60	147 (39.2)	228 (60.8)		101 (32.9)	206 (67.1)		44 (69.8)	19 (30.2)	
Menopausal Status									
Pre-	215 (48.6)	227 (51.4)	0.222 (1.53)	106 (35.2)	195 (64.8)	0.377 (0.78)	107 (78.7)	29 (21.3)	0.115 (2.49)
Post-	307 (44.9)	377 (55.1)		206 (38.3)	332 (61.7)		100 (70.4)	42 (29.6)	
Tumour Size (cm)									
≤2.0	226 (41.3)	321 (58.7)	**0.001** (11.10)	149 (33.8)	292 (66.2)	**0.045** (4.01)	73 (74.5)	25 (25.5)	0.885 (0.02)
> 2.0	300 (51.2)	286 (48.8)		163 (40.4)	240 (59.6)		137 (75.3)	45 (24.7)	
Stage									
1	318 (46.2)	371 (53.8)		188 (36.4)	328 (63.6)		126 (76.4)	39 (23.6)	
2	154 (44.9)	189 (55.1)	0.354 (2.07)	96 (36.6)	166 (63.4)	0.805 (0.44)	58 (72.5)	22 (27.5)	0.794 (0.46)
3	52 (53.1)	46 (46.9)		26 (40.6)	38 (59.4)		26 (76.5)	8 (23.5)	
Tumour Type									
Ductal NST	350 (52.2)	321 (47.8)		173 (39.8)	262 (60.2)		177 (74.4)	59 (25.3)	
Lobular	30 (27.8)	78 (72.2)		30 (28.6)	75 (71.4)		0 (0.0)	2 (100)	
Tubular mixed	67 (34.2)	129 (65.8)		62 (33.2)	125 (66.8)		5 (62.5)	3 (37.5)	
Medullary-like	23 (92.0)	2 (8.0)	**<0.001** (72.35)	2 (66.7)	1 (33.3)	**0.002** (20.53)	21 (95.5)	1 (4.5)	**0.006** (18. 08)
Special types *	10 (20.0)	40 (80.0)		8 (17.8)	37 (82.2)		1 (50.0)	1 (50.0)	
Mixed NST and Lobular	19 (43.2)	25 (56.8)		18 (43.9)	23 (56.1)		1 (33.3)	2 (66.7)	
Mixed NST and other special type	12 (60.0)	8 (40.0)		12 (63.2)	7 (38.6)		0 (0.0)	1 (100)	
Grade									
1	53 (29.1)	129 (70.9)		50 (29.1)	122 (70.9)		1 (20.0)	4 (80.0)	
2	120 (32.1)	254 (67.9)	**<0.001** (102.88)	107 (30.6)	243 (69.4)	**<0.001** (26.94)	12 (57.1)	9 (42.9)	**0.002** (12.83)
3	351 (61.1)	223 (38.9)		153 (47.8)	167 (52.2)		197 (77.9)	56 (22.1)	
Tubules									
1	19 (29.2)	46 (70.8)		18 (30.0)	42 (70.0)		0 (0.0)	3 (100.0)	
2	135 (38.5)	216 (61.5)	**<0.001** (23.60)	107 (34.6)	202 (65.4)	0.318 (2.29)	28 (70.0)	12 (30.0)	**0.007** (9.95)
3	352 (51.5)	332 (48.5)		173 (38.4)	277 (61.6)		179 (76.8)	54 (23.2)	
Pleomorphism									
1	12 (41.4)	17 (58.6)		11 (44.0)	14 (56.0)		0 (0.0)	2 (100.0)	
2	124 (29.9)	291 (70.1)	**<0.001** (71.05)	115 (29.0)	281 (71.0)	**<0.001** (17.21)	9 (52.9)	8 (47.1)	**0.004** (10.91)
3	367 (56.2)	286 (43.8)		170 (42.9)	226 (57.1)		197 (77.0)	59 (23.0)	
Mitosis									
1	107 (27.9)	276 (72.1)		101 (27.6)	265 (72.4)		5 (38.5)	8 (61.5)	
2	82 (39.8)	124 (60.2)	**<0.001** (106.38)	64 (36.6)	111 (63.4)	**<0.001** (27.98)	18 (60.0)	12 (40.0)	**0.001** (14.82)
3	317 (62.0)	194 (38.0)		133 (47.8)	145 (52.2)		184 (79.0)	49 (21.0)	
LVI									
Negative	340 (46.0)	399 (54.0)	0.774 (0.08)	202 (36.5)	351 (63.5)	0.753 (0.10)	135 (75.4)	44 (24.6)	0.793 (0.07)
Definite	182 (46.9)	206 (53.1)		108 (37.6)	179 (62.4)		74 (74.0)	26 (26.0)	
NPI									
GPG	102 (30.4)	234 (69.6)		92 (29.4)	221 (70.6)		7 (46.7)	8 (53.3)	
MPG	313 (51.7)	292 (48.3)	**<0.001** (51.75)	167 (40.0)	250 (60.0)	**0.001** (13.48)	145 (77.5)	42 (22.5)	**0.029** (7.08)
PPG	111 (57.8)	81 (42.2)		53 (46.5)	61 (53.5)		58 (74.4)	20 (25.6)	

NPI: Nottingham Prognostic Index, GPG; Good Prognostic Group; MPG: Moderate Prognostic Group; PPG: Poor Prognostic Group; LVI: Lympho-Vascular Invasion

*: include invasive mucinous, invasive cribriform invasive tubular and invasive papillary carcinomas.

Table 2: Association between AR and the expression other markers in the whole series and in ER-positive BC

	AR Expression in the whole series			AR Expression in the ER-Positive tumours			AR Expression in the ER-Negative tumours		
	Negative/low N (%)	Positive/high N (%)	p-value (χ^2)	Negative/low N (%)	Positive/high N (%)	p-value (χ^2)	Negative/low N (%)	Positive/high N (%)	p-value (χ^2)
ER				-	-	-	-	-	-
Negative	210 (74.7)	71 (25.3)	**<0.001** (120.77)						
Positive	313 (37.0)	533 (63.0)							
PgR							-	-	-
Negative	280 (64.4)	155 (35.6)	**<0.001** (92.00)	78 (47.0)	88 (53.0)	**0.004** (8.48)			
Positive	229 (34.8)	429 (65.2)		229 (34.8)	429 (65.2)				
HER2									
Negative	425 (44.8)	523 (55.2)	**0.003** (8.76)	269 (36.0)	479 (64.0)	**0.034** (4.50)	153 (79.3)	40 (20.7)	0.061 (3.50)
Positive	82 (58.2)	59 (41.8)		35 (48.6)	37 (51.4)		47 (68.1)	22 (31.9)	
BC Molecular Class									
Luminal	279 (35.9)	499 (64.1)	**<0.001 (131.69)**	278 (35.8)	498 (64.2)	**0.034** (4.63)	-	-	**0.049** (7.85)
HER2-positive	83 (58.5)	59 (41.5)		35 (48.6)	37 (51.4)		48 (68.6)	22 (31.4)	
TN Non-Basal	43 (89.6)	5 (10.4)		-	-		43 (89.6)	5 (10.4)	
TN Basal-like	109 (76.8)	33 (23.2)		-	-		109 (76.80	33 (23.2)	
AGTR1									
Negative	64 (43.8)	82 (56.2)		30 (28.6)	75 (71.4)		34 (82.9)	7 (17.1)	0.133 (4.03)
Low	77 (41.2)	110 (58.8)	**0.004** (11.16)	51 (34.7)	96 (65.3)	**0.014** (8.50)	26 (66.7)	13 (33.3)	
Positive	167 (55.5)	134 (44.5)		94 (44.5)	117 (55.5)		73 (81.1)	17 (18.9)	
PELP1									
Negative	46 (36.8)	79 (63.2)		28 (28.0)	72 (72.0)		18 (75.0)	6 (25.0)	0.91 (0.19)
Low	245 (48.9)	256 (51.1)	**0.049** (6.05)	152 (39.9)	229 (60.1)	**0.026** (7.27)	93 (78.2)	26 (21.8)	
High	58 (44.6)	72 (55.4)		25 (28.7)	62 (71.3)		33 (76.7)	10 (23.3)	
CARM1									
Negative	102 (52.8)	91 (47.2)		77 (47.2)	86 (52.8)		24 (85.7)	4 (14.3)	0.292 (2.46)
Low	168 (43.9)	215 (56.1)	0.078 (5.11)	97 (33.9)	189 (66.1)	**0.008** (9.59)	71 (74.0)	25 (26.0)	
Positive	79 (51.3)	75 (48.7)		30 (31.6)	65 (68.4)		50 (82.0)	11 (18.0)	
CD71									
Negative	128 (39.0)	200 (61.0)	**<0.001** (20.60)	93 (33.2)	187 (66.8)	**0.016** (5.82)	35 (74.5)	12 (25.5)	0.099 (4.63)
Positive	244 (55.6)	195 (44.4)		124 (43.1)	164 (56.9)		119 (80.4)	29 (19.6)	
Cyclin B1									
Negative	179 (47.1)	201 (52.9)	0.983 (0.00)	99 (35.7)	178 (64.3)	0.775 (0.08)	80 (78.4)	22 (21.6)	0.938 (0.13)
Positive	126 (47.2)	141 (52.8)		70 (37.0)	119 (63.0)		56 (71.8)	22 (28.2)	
FOXa1						**<0.001** (27.78)			
Negative	255 (62.0)	156 (38.0)	**<0.001** (69.40)	124 (49.8)	125 (50.2)		131 (80.9)	31 (19.1)	0.250 (2.78)
Positive	104 (31.3)	228 (68.7)		82 (27.8)	213 (72.2)		22 (61.1)	14 (38.9)	
GATA3						**<0.001** (12.34)			
Negative/Low	280 (56.9)	212 (43.1)	**<0.001** (39.42)	144 (45.0)	176 (55.0)		136 (79.1)	36 (20.9)	0.250 (0.76)
Positive	43 (27.9)	111 (72.1)		42 (28.0)	108 (72.0)		0 (0.0)	1 (100.0)	
P53						0.493 (0.47)			
Negative	331 (42.5)	448 (57.5)	**<0.001** (14.36)	235 (36.0)	418 (64.0)		95 (76.6)	29 (23.4)	0.305 (1.05)
Positive	168 (55.3)	136 (44.7)		63 (68.9)	99 (61.1)		105 (73.9)	37 (26.1)	
TK1						0.219 (1.51)			
Negative	109 (38.0)	178 (62.0)	**0.004** (8.29)	78 (31.2)	172 (68.8)		31 (86.1)	5 (13.9)	**0.011** (6.54)
Positive	162 (49.5)	165 (50.5))		78 (36.6)	135 (63.4)		84 (73.7)	30 (26.3)	
BEX1						**<0.001** (24.10)			
Negative/Low	144 (60.0)	96 (40.0)	**<0.001** (19.28)	91 (53.5)	79 (46.5)		53 (76.8)	16 (23.2)	0.055 (3.70)
Positive	202 (42.6)	272 (57.4)		112 (31.3)	246 (68.7)		89 (78.1)	25 (21.9)	
RERG						**0.002** (9.74)			
Negative	282 (49.8)	284 (50.2)	**<0.001** (15.92)	162 (38.8)	255 (61.2)		120 (81.6)	27 (18.4)	0.615 (0.25)
Positive	63 (33.2)	127 (66.8)		37 (24.7)	113 (75.3)		26 (65.0)	14 (35.0)	

Ki67						**0.029** (4.76)			
Negative	126 (34.5)	239 (65.5)	**<0.001** (30.49)	107 (32.3)	224 (67.7)		19 (61.3)	12 (38.7)	0.124 (2.36)
Positive	292 (53.1)	258 (46.9)		144 (40.3)	213 (59.7)		147 (76.6)	45 (23.4)	

Table 3: Cox proportional hazard analysis for predictors of BCSS within the whole series and ER+ tumours for the expression AR and other co-variates:

Variable	In the whole series			In ER positive tumours		
	p value	**HR**	**95% CI**	**p value**	**HR**	**95% CI**
AR	**0.033**	0.80	0.64-0.98	**0.007**	0.71	0.56-0.91
Tumour size	**0.011**	1.33	1.07-1.66	**0.003**	1.49	1.15-1.94
Nodal Stage	**<0.001**	1.80	1.55-2.08	**<0.001**	1.62	1.35-1.94
Tumour grade	**<0.001**	1.61	1.36-1.91	**<0.001**	1.72	1.42-2.07

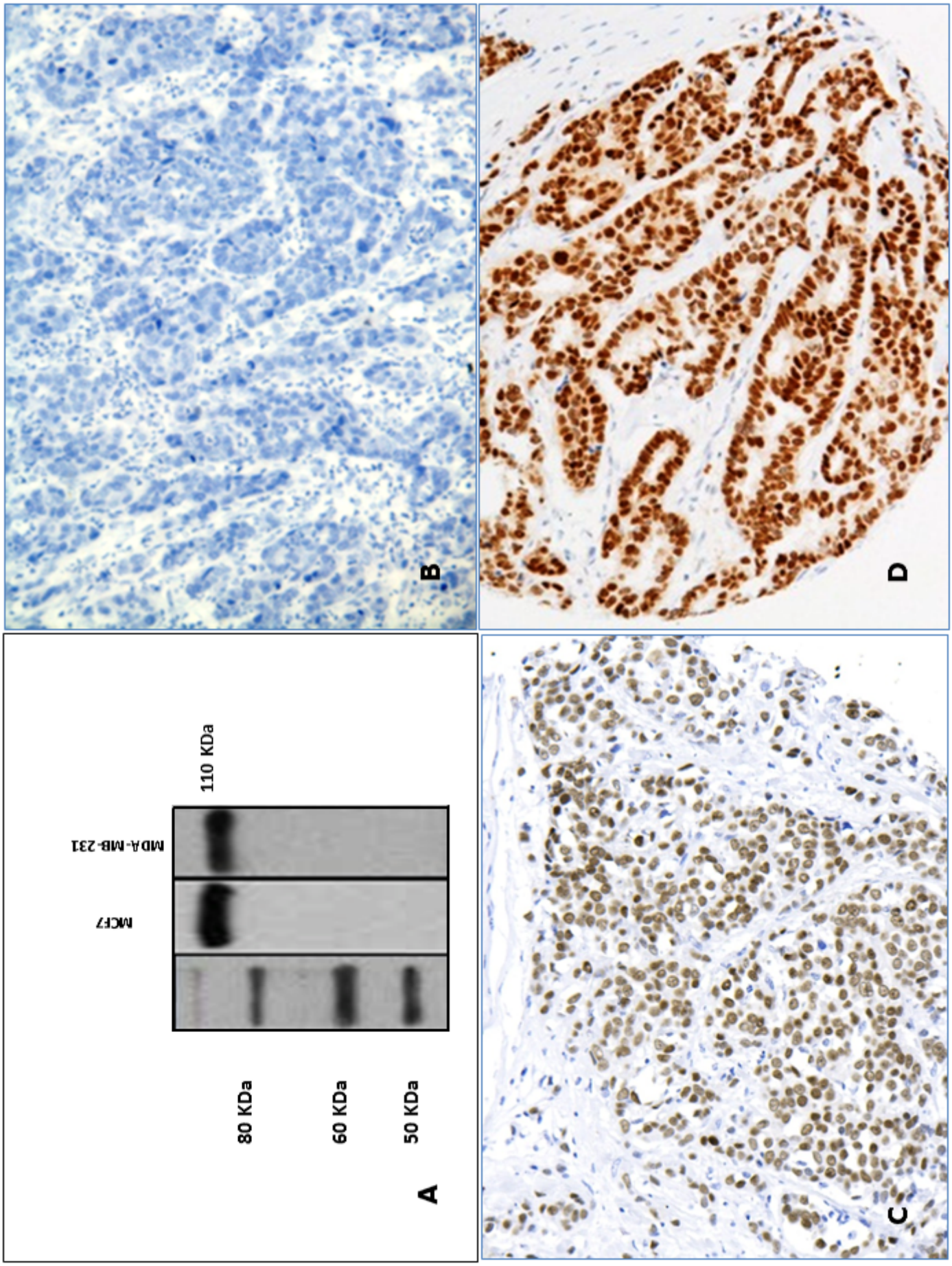


Figure 1

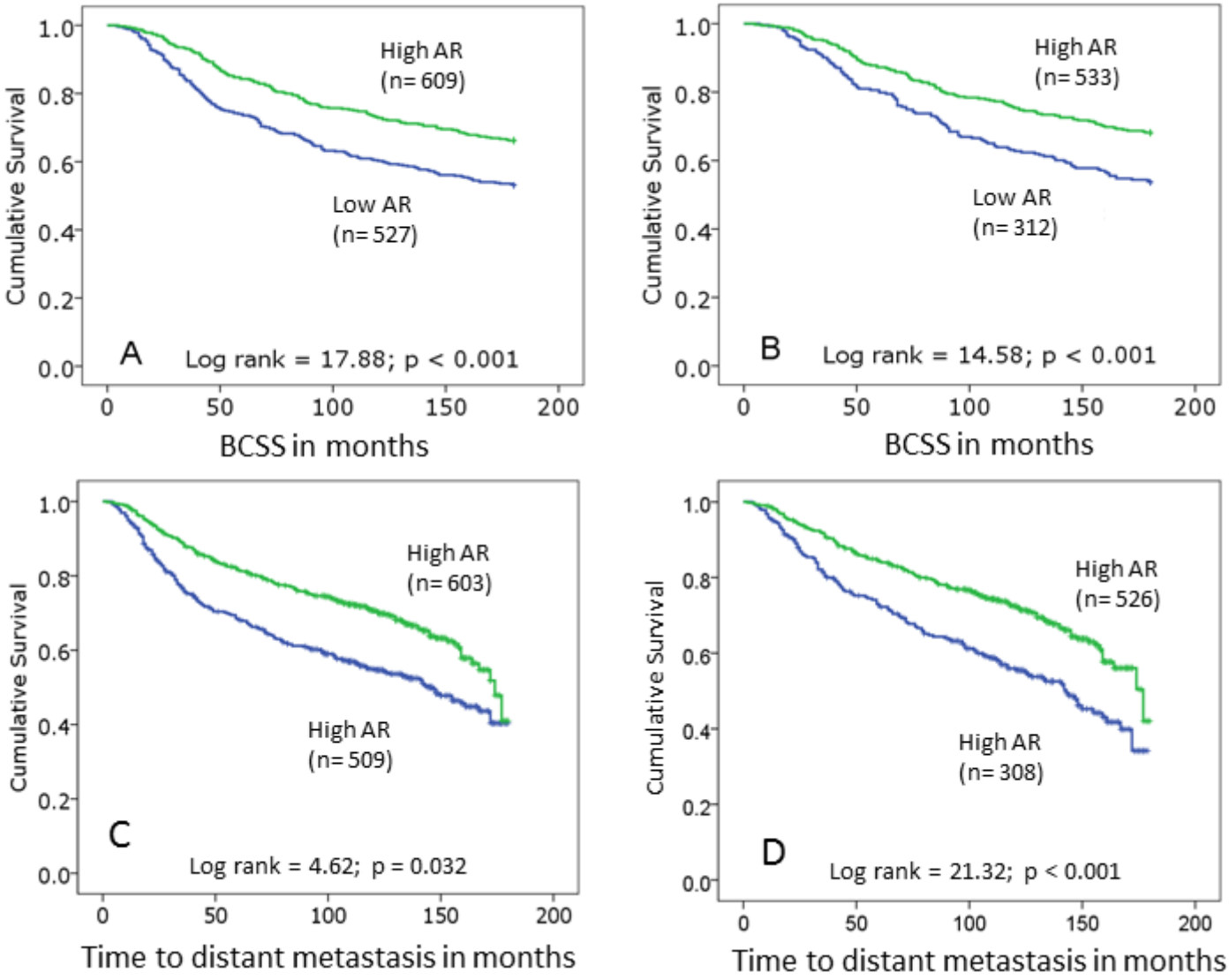

Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
High AR
(n= 609)
Low AR
(n= 527)
A
Log rank = 17.88; p < 0.001
0
50
100
150
200
BCSS in months
High AR
(n= 533)
Low AR
(n= 312)
B
Log rank = 14.58; p < 0.001
BCSS in months
High AR
(n= 603)
High AR
(n= 509)
C
Log rank = 4.62; p = 0.032
Time to distant metastasis in months
High AR
(n= 526)
High AR
(n= 308)
D
Log rank = 21.32; p < 0.001
Time to distant metastasis in months

Figure 2

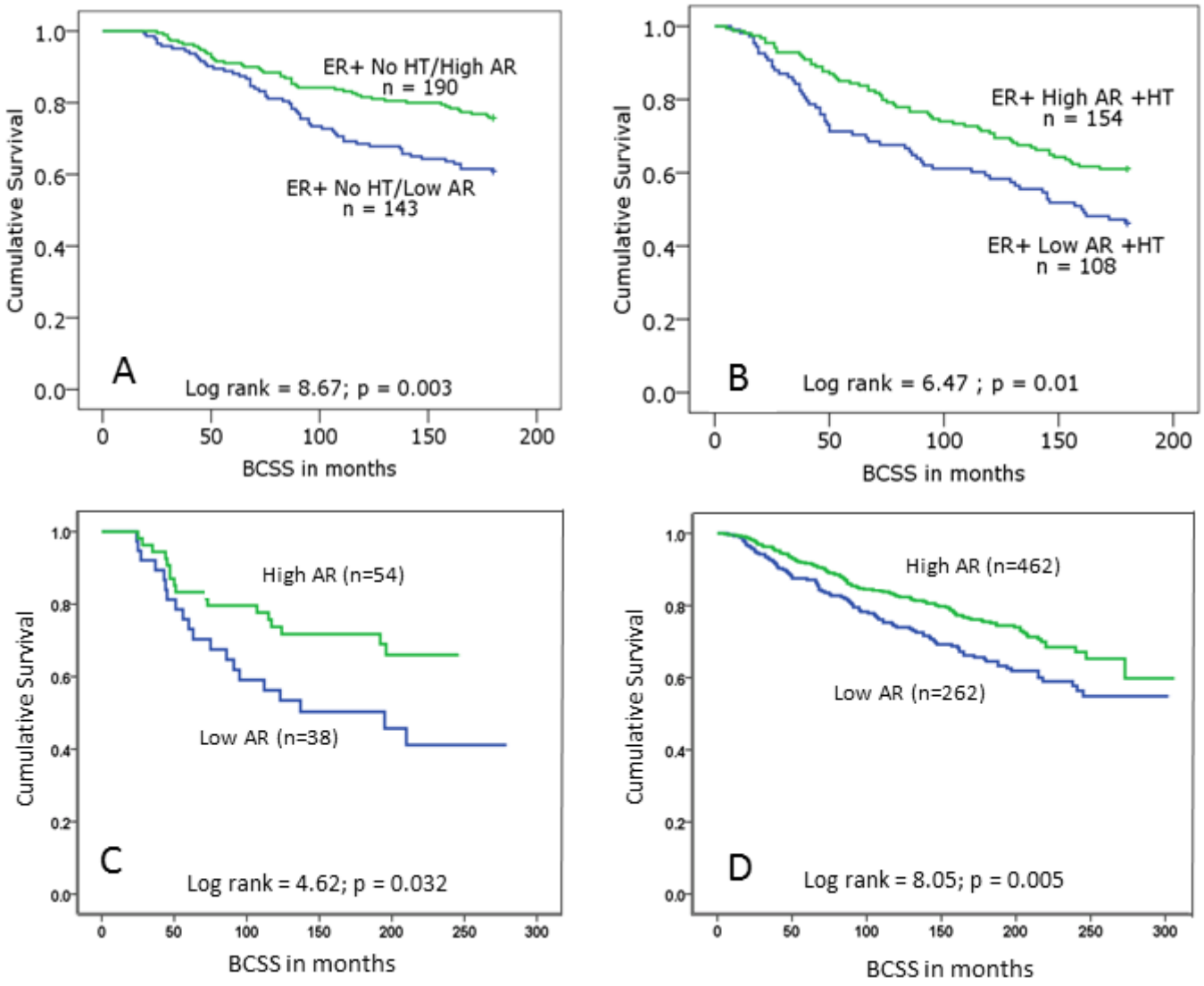

Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
ER+ No HT/High AR
n = 190
ER+ No HT/Low AR
n = 143
A
Log rank = 8.67; p = 0.003
0
50
100
150
200
BCSS in months
ER+ High AR +HT
n = 154
ER+ Low AR +HT
n = 108
B
Log rank = 6.47 ; p = 0.01
High AR (n=54)
Low AR (n=38)
C
Log rank = 4.62; p = 0.032
250
300
High AR (n=462)
Low AR (n=262)
D
Log rank = 8.05; p = 0.005

Figure 3

Supplementary Table 1: Clinico-pathological features of the study cohort:

Clinicopathological characteristics	N (%)*
Age	
≤50	416 (41.9)
>50	578 (58.1)
Missing	147
Menopausal Status	
Pre-menopausal	410 (41.6)
Post- menopausal	576 (58.4)
Missing	155
Tumour Size (cm)	
≤2.0	474 (47.8)
>2.0	518 (52.2)
Missing	149
Tumour Type	
Ductal NST	671 (60.2)
Lobular	108 (9.7)
Tubular mixed	196 (17.6)
Medullary-like	25 (2.2)
Special types	50 (4.5)
Mixed NST and Lobular	44 (3.9)
Mixed NST and other special type	20 (1.8)
Missing	27
NPI**	
GPG	284 (28.6)
MPG	535 (53.9)
PPG	173 (17.4)
Missing	149
Stage	
1	603 (60.9)
2	306 (30.9)
3	81 (8.2)
Missing	151
Grade	
1	153 (15.5)
2	324 (32.7)
3	513 (51.8)
Missing	151
LVI	
Negative	635 (64.3)
Definite	352 (35.7)
Missing	154
Distant metastases	
No	626 (63.3)
Yes	363 (36.7)
Missing	152
BC Molecular classes	

Luminal	778 (68.2)
HER2 positive	142 (12.4)
Triple negative (TN) non-Basal	48 (4.2)
TN Basal-like	142 (12.4)
Missing	31

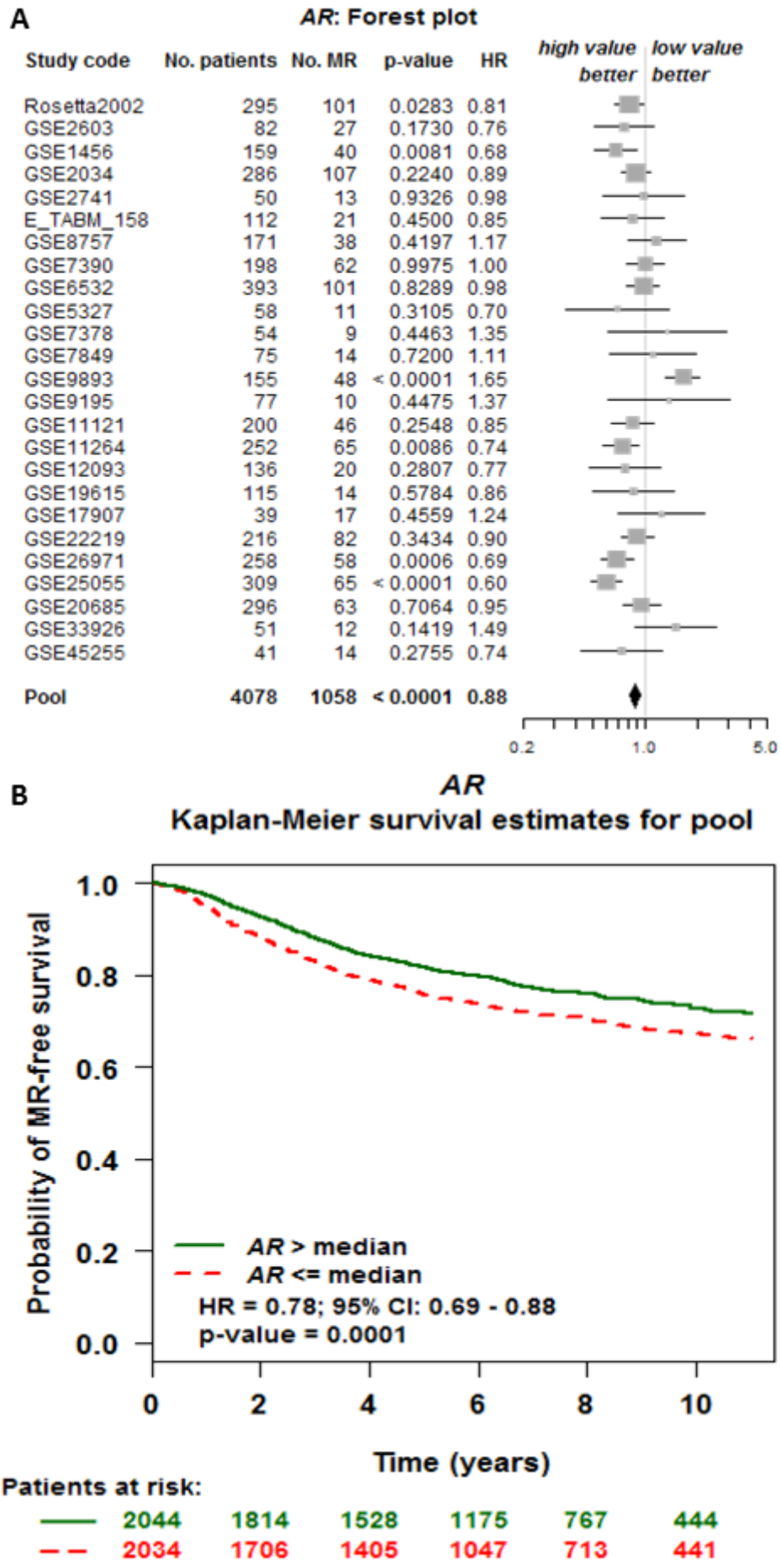
* These are the valid percentages (excluding the missing cases for each parameter).

** NPI: Nottingham Prognostic Index, GPG; Good Prognostic Group; MPG: Moderate Prognostic Group; PPG: Poor Prognostic Group; LVI: Lympho-Vascular Invasion.

Supplementary Table 2: The prognostic impact *AR* mRNA in the external validation cohorts using bc-GenExMiner v4.0 for publicly available online datasets (n=25 datasets, 4078 patients).

Cohort	Reference	p-value	HR	95% CI	Number of patients	Number of Metastatic Recurrence
Rosetta2002	Van de Vijver *et al.*, 2002	**0.0283***	0.81	0.67 - 0.98	295	101
GSE2603	Minn *et al.*, 2005	0.1730	0.76	0.51 - 1.13	82	27
GSE1456	Pawitan *et al.*, 2005	**0.0081***	0.68	0.51 - 0.90	159	40
GSE2034	Wang *et al.*, 2005	0.2240	0.89	0.73 - 1.08	286	107
GSE2741	Weigelt *et al.*, 2005	0.9326	0.98	0.57 - 1.67	50	13
E_TABM_158	Chin *et al.*, 2006	0.4500	0.85	0.56 - 1.29	112	21
GSE8757	Chin *et al.*, 2007	0.4197	1.17	0.80 - 1.72	171	38
GSE7390	Desmedt *et al.*, 2007	0.9975	1.00	0.78 - 1.28	198	62
GSE6532	Loi *et al.*, 2007	0.8289	0.98	0.79 - 1.20	393	101
GSE5327	Minn *et al.*, 2007	0.3105	0.70	0.35 - 1.39	58	11
GSE7378	Zhou *et al.*, 2007	0.4463	1.35	0.62 - 2.96	54	9
GSE7849	Anders *et al.*, 2008	0.7200	1.11	0.63 - 1.97	75	14
GSE9893	Chanrion *et al.*, 2008	**< 0.0001****	1.65	1.32 - 2.07	155	48
GSE9195	Loi *et al.*, 2008	0.4475	1.37	0.61 - 3.06	77	10
GSE11121	Schmidt *et al.*, 2008	0.2548	0.85	0.65 - 1.12	200	46
GSE11264	Jézéquel *et al.*, 2009	**0.0086***	0.74	0.58 - 0.92	252	65
GSE12093	Zhang *et al.*, 2009	0.2807	0.77	0.48 - 1.24	136	20
GSE19615	Li *et al.*, 2010	0.5784	0.86	0.50 - 1.47	115	14
GSE17907	Sircoulomb *et al.*, 2010	0.4559	1.24	0.70 - 2.21	39	17
GSE22219	Buffa *et al.*, 2011	0.3434	0.90	0.73 - 1.12	216	82
GSE26971	Filipits *et al.*, 2011	**0.0006***	0.69	0.56 - 0.86	258	58
GSE25055	Hatzis *et al.*, 2011	**< 0.0001***	0.60	0.49 - 0.74	309	65
GSE20685	Kao *et al.*, 2011	0.7064	0.95	0.75 - 1.22	296	63
GSE33926	Kuo *et al.*, 2012	0.1419	1.49	0.87 - 2.55	51	12
GSE45255	Nagalla *et al.*, 2013	0.2755	0.74	0.43 - 1.27	41	14
Pooled Data		**< 0.0001***	**0.88**	**0.83 - 0.94**	**4078**	**1058**

*: High *AR* mRNA associated is good prognostic, **: Low *AR* mRNA is poor prognostic. HR: Hazard Ratio, CI: confidence Interval.



Supplementary Figure 1