Ubiquitin-conjugating enzyme 2C (UBE2C) is a poor prognostic biomarker in invasive breast cancer

Yousif Kariri[1, 2], Michael S. Toss[1], Mansour Alsaleem[1, 3], Khloud A. Elsharawy[1, 4], Chitra Joseph[1], Nigel P. Mongan[5, 6], Andrew R. Green[1], Emad A. Rakha[1].

[1]Academic Unit for Translational Medical Sciences, School of Medicine, University of Nottingham Biodiscovery Institute, University Park, Nottingham, NG7 2RD, UK.

[2] Department of Clinical Laboratory Science, Faculty of Applied Medical Science, Shaqra University 33, Shaqra 11961, Saudi Arabia.

[3] Department of Applied Medical Science, Applied Collage in Unazyzah, Qassim University, Qassim, Saudi Arabia.

[4] Department of Zoology, Faculty of Science, Damietta University, Damietta, Egypt, 34517.

[5] Biodiscovery Institute, Faculty of Medicine and Health Sciences, School of Veterinary Medicine and Science, University of Nottingham, NG7 2RD, UK

[6] Department of Pharmacology, Weill Cornell Medicine, New York, NY, 10065, USA.

Correspondence:

Professor Emad Rakha

Department of Histopathology, Nottingham University Hospital NHS Trust,

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

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

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

Keywords: UBE2C, breast cancer, lymphovascualr invasion, progression, prognosis, outcome.

Running title: Expression of UBE2C in invasive breast cancer.

ABSTRACT:

Background: The Ubiquitin-conjugating enzyme 2C (UBE2C) is essential for the ubiquitin-proteasome system and is involved in cancer cell migration and apoptosis. This study aimed to determine the prognostic value of UBE2C in invasive breast cancer (BC).

Methods: *UBE2C* was evaluated using the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC; n=1980), The Cancer Genome Atlas (TCGA; n=854) and Kaplan-Meier Plotter (n=3951) cohorts. UBE2C protein expression was assessed using immunohistochemistry in the BC cohort (n=619). The correlation between UBE2C, clinicopathological parameters and patient outcome was assessed.

Results: High *UBE2C* mRNA and protein expressions were correlated with features of poor prognosis, including high tumour grade, large size, the presence of lymphovascular invasion (LVI), hormone receptor negativity and HER2 positivity. High *UBE2C* mRNA expression showed a negative association with E-cadherin, and a positive association with adhesion molecule N-cadherin, matrix metalloproteinases and cyclin related genes. There was a positive correlation between high UBE2C protein expression and cell cycle associated biomarkers, p53, Ki67, EGFR and PI3K. High UBE2C protein expression was an independent predictor of poor outcome (p=0.011, HR =1.45, 95% CI; 1.10-1.93).

Conclusion: This study indicates that UBE2C is an independent prognostic biomarker in BC. These results warrant further functional validation for UBE2C as a potential therapeutic target in BC.

INTRODUCTION

Breast cancer (BC) is a heterogeneous disease comprising several biological subtypes and shows diverse behaviours and responses to therapy [1].In-depth investigation of the transcriptomic and proteomic expression of the underlying genetic pathways which contribute to both invasion and metastasis can be critical to decipher the complex molecular make up of BC and refine and improve its clinical management.

The ubiquitination process is an essential protein degradation mechanism that serves to protect cellular integrity by degrading abnormal and short-life proteins. Moreover, it contributes to the cellular processes that induce cell cycle progression, transcription and apoptosis [2]. Ubiquitin-conjugating enzyme 2C (UBE2C) is a participant in the ubiquitin-conjugating enzyme complex, and it also plays an essential role in the ubiquitin-proteasome system, which normally regulates key checkpoints in the cell cycle via targeting the cell cycle regulators [3]. The UBE2C-encoded protein is involved in mitotic cyclin destructions and cell cycle progression; hence, it potentially could participate in cancer development [4]. Previous studies have identified high UBE2C expression in several types of cancer, including head and neck squamous cell carcinoma [5], gastrointestinal [6] and endometrial cancer [7].

Lymphovascular invasion (LVI), which is indicated by the presence of tumour cells within lymphatic vessels, is considered one of the prerequisites for BC metastasis [8-10]. However, the key molecular processes associated with BC-LVI progression remain poorly understood. Hence, further investigations are required to detect both biological and molecular mechanisms underlying LVI. The results of such investigations should prove vital in terms of developing targeted treatment strategies that can help in improving patient outcomes. Although several prior studies have reported that high expression of UBE2C plays a major role in the progression of BC [11-15], but its role in BC-LVI remains unclear. Based on the findings of the aforementioned studies, we hypothesised that UBE2C plays a significant role in BC progression and metastasis. Here we investigated the expression of UBE2C in BC at both the transcriptomic and proteomic levels to determine its association with various clinicopathological features including LVI, other related genes and patient outcomes using several well-characterised BC cohorts and datasets.

MATERIAL AND METHODS

Study cohorts

To investigate the prognostic significance of *UBE2C* mRNA expression in BC, gene expression data were obtained from the TNM-plot (https://www.tnmplot.com/) and UALCAN (http://ualcan.path.uab.edu/index.html) datasets, which together include 1097 primary, 7 metastatic tumours and 113 normal tissue samples [16,17]. Likewise, both the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (n=1980) [18] and The Cancer Genome Atlas (TCGA) (n=854) [19] datasets were used as discovery cohorts to assess and explore the prognostic value of *UBE2C* expression at the genomic level. To validate the prognostic value of *UBE2C* mRNA expression, the Kaplan Meier (KM) Plotter (n=3951) online data set (https://kmplot.com/analysis/) [20], was used.

UBE2C protein expression was measured by immunohistochemistry (IHC) in a large BC cohort (n=619) with detailed clinical information comprising patients presented at Nottingham City Hospital, Nottingham, United Kingdom as previously described [21]. For management purposes, Nottingham Prognostic Index (NPI) and Oestrogen Receptor (ER) status were used to classify patients into clinically relevant groups. Patients with a good prognostic NPI score (≤ 3.4) received no adjuvant therapy, whereas patients with poor prognostic NPI score (>3.4) received endocrine treatment if ER status was positive and received chemotherapy (classical cyclophosphamide, methotrexate and 5-fluorouracil (CMF)) if ER status was negative. None of the patients in this study received neoadjuvant therapy or anti-human epidermal growth factor receptor 2 (HER2) targeted therapy. The clinicopathological features for the cohort series were summarised previously [21,22].

To investigate the interactions between UBE2C expression and other related biomarkers, previous available data [23-25] has been used. This includes DNA and cell cycle regulator (p53, CDCA5), proliferation marker (Ki67), adhesion molecules (E-cadherin (CDH1) and N-cadherin (CDH2), basal-phenotype (CK5 and CK14 positive), phosphatidylinositol 3-kinase (PI3K) and epidermal growth factor receptor (EGFR)

UBE2C protein expression evaluation

Prior to IHC staining, the validity of the primary UBE2C antibody (WHO0011065M1, Sigma-Aldrich, Gillingham, UK, 1:300) was checked using immunoblotting. The specificity of the UBE2C was validated in SKBR3 human BC cells (obtained from the American Type Culture Collection, Rockville, MD, USA). The rabbit β-actin antibody (A5441, clone AC-15, Sigma-Aldrich, Gillingham, UK) was used at 1:5000 as a housekeeping protein and showed a band at approximately 42 KDa. A single specific band for the UBE2C protein was detected at the expected molecular weight of ~20 KDa after incubation overnight (Supplementary Figure 1A).

Fourteen full face sections of BC cases, representative of several molecular subtypes and tumour grade, were used to evaluate the distribution of UBE2C expression. Patients' samples were arrayed into tumour microarrays (TMA) as previously described [26]. Citrate antigen retrieval (pH 6.0) was used, and samples were incubated overnight at 4 °C with UBE2C antibody diluted (1:100). Novolink Max Polymer Detection kit (Leica, Newcastle, UK) was used to express the immunoreactivity of UBE2C [21]. UBE2C stained slides were scanned using high resolution digital images (NanoZoomer; Hamamatsu Photonics, Welwyn Garden City, UK) at 20x magnification and visualised on viewing software (Xplore; Philips, UK) to assess the protein expression level. A semi-quantitative evaluation was used to assess a modified histochemical score (H-score) [27] which is combined with the staining intensity (0-3) multiplied by the proportion of tumour cells (0-100). The staining intensity was categorised into four groups: 3 (strong staining); 2 (moderate staining); 1 (weak staining) and 0 (no staining). The final H-score was obtained by giving a range of 0 to 300. Cores with less than 15% tumour areas and/or with folded tissue were not assessed. The interobserver concordance was checked by doing a blind double scoring for two researchers (YK and SA).

Statistical analysis

The data analysis was presented using SPSS statistical software (IBM SPSS Statistic, Version 24.0, Chicago, IL, USA). The mRNA and protein expressions were categorised into low and high subgroups according to their median (METABRIC; 9.13, TCGA; 533, protein; 20 H-score) cut-off. Inter-observer agreement in UBE2C IHC scoring was evaluated using intra-class correlation coefficient (ICC). The

associations between mRNA expression of *UBE2C* and adhesion molecules, metalloproteinase (*MMPs*), cyclin and cell cycle related genes were analysed by using Person's correlation test. The Chi square test was used to study the correlation between UBE2C expression and the other categorical variables in both transcriptomic and proteomic levels. Kaplan-Meier survival test was performed to assess the correlation with patients' outcome. Cox regression model was used for multivariate analysis. *P* value of <0.05 was used to detect the statistical significance.

This study followed the reporting recommendations for tumour markers prognostic studies (REMARK) criteria [28].

RESULTS

Transcriptomic and Genomic expression of *UBE2C*

In both the TNM plotter and ULACAN datasets, high *UBE2C* mRNA expression was identified more in BC when compared with the normal breast tissues (Supplementary Figure 1B). Among the different molecular subtypes, the expression of *UBE2C* was higher in the HER2-enriched BC and triple negative (TNBC) than in the luminal-A class (Supplementary Figure 1C and Table 1). High *UBE2C* mRNA expression was significantly associated with the presence of LVI (METABRIC cohort: p=0.002, TCGA cohort: p<0.001) and other factors characteristics of a poor prognosis, including larger tumour size (p< 0.001), high tumour grade (p< 0.001), ER and progesterone receptor (PR) negativity (p<0.001) and HER2 positivity (p<0.001; Table 1). High *UBE2C* expression was also associated with a high nodal stage in the METABRIC cohort (p<0.001) (Table 1).

UBE2C mRNA expression and related biomarkers

In the METABRIC cohort, high *UBE2C* mRNA expression showed an association with epithelial-mesenchymal transition (EMT) phenotype, specifically negative correlation with *CDH1* and positive association with *CDH2* (p<0.001) (Table 2). High *UBE2C* mRNA expression also showed a strong positive association with several members of the *MMPs* family (*MMP7, MMP9, MMP12, MMP14, MMP15, MMP20, MMP21* and *MMP25*), proliferation related genes (*CDK1, CDK2, CDK4, CDK5,*

CDK6, CDKN2A, CCNB1, CCNE1, CCNE2, CCNA1, CCNA2, CCNB2 and *CCND3*) and cell cycle related genes (*CDCA5* and *CDC20*) in both METABRIC and TCGA datasets (all p<0.05; Table2).

UBE2C mRNA expression and patients' outcome

High *UBE2C* mRNA expression was significantly associated with shorter BC specific survival (BCSS) in the METABRIC cohort (p<0.001, HR = 2.50, 95% CI 2.07–3.01; Figure 1A), in the TCGA cohort (p=0.006, HR = 2.41, 95% CI 2.01–2.90; Figure 1B) and in the KM-Plotter BC online data sets (p<0.001, HR = 1.76, 95% CI 1.57–1.96; Figure 1C). Multivariate analysis in METABRIC cohort observed that *UBE2C* expression was an independent prognostic marker significantly associated with poor patient outcome in terms of BCSS (p<0.001, HR 1.90, 95% CI; 1.50-2.38), regardless of LVI, tumour size, ER and HER2 status (Table 3).

Categorisation of the transcriptomic cohorts based on the LVI status showed that high *UBE2C* mRNA expression was strongly associated with poor patient outcome in the LVI positive BC in both the METABRIC cohort (p<0.001, HR = 2.10, 95% CI; 1.70–2.53; Figure 1D) and the TCGA cohort (p=0.001, HR = 2.10, 95% CI; 1.40–3.17; Figure 1E). Furthermore, high *UBE2C* mRNA expression showed a non-significant association with the LVI negative BC in the METABRIC cohort (p=0.221, HR = 1.43, 95% CI; 0.80–2.60; Supplementary Figure 2A) and the TCGA cohort (p=0.537, HR 1.21, 95% CI; 0.65-2.26; Supplementary Figure 2B).

UBE2C protein expression

Full face sections of BC showed even distribution for UBE2C protein expression, which indicated the suitability of TMA to assess UBE2C protein expression. UBE2C protein expression was detected prominently in the cytoplasm of invasive tumour cells. Following double scoring of cases, a good concordance rate was obtained between the two the observers (ICC= 0.7, p= 0.024). Therefore, the main observer (YA) scoring was considered in the final analysis. The distribution of UBE2C protein expression showed a range from absent to high (H-score 0-160), and for dichotomisation into negative/low and high expression, the median H-score 20 was used. 376 (61%) of cases showed low expression, whereas 243 (39%) cases with high expression (Figure 2B and 2C).

High expression of UBE2C was significantly associated with the presence of LVI (p=0.009), and other variables of poor prognosis including the presence of nodal status, high tumour grading, larger tumour size, poor NPI, lack of ER and PR receptors expression, and HER2 positivity (Table 4). When we stratified the protein expression based on BC histological subtypes, high UBE2C protein expression was strongly associated with ductal NST BC tumour compared to other types (p<0.001; Table 4).

High UBE2C protein expression was strongly correlated with high p53 expression (p<0.001), high Ki67 index (p=0.008), basal-phenotype biomarkers (p=0.002), EGFR (p=0.003), N-cadherin (p=0.033), stromal immune markers CD8 and CD68 (all: p<0.001), cyclin B (p=0.041), and high level of PI3K (p=0.019; Table 4). Among BC IHC subtypes, high UBE2C protein was indicated to be obtained more with HER2-enriched and TNBC subtype compared to other subtypes (p<0.001; Table 4).

Patients who had high UBE2C protein expression displayed poor BCSS (p=0.011, HR =1.45, 95% CI; 1.10-1.93; Figure 3A) compared to patients who had low expression. Moreover, patients with high UBE2C protein expression showed a significant poor 10 years BC disease free survival (BCDFS) (p=0.019, HR= 1.43, 95% CI; 1.06-1.91; Figure 3B). Multivariate analysis revealed that UBE2C expression associated with poor patients' outcome in term of BCSS (p=0.013, HR = 1.60, 95% CI; 1.10-2.30), independent on other prognostic parameters including LVI, tumour size, ER and HER2 status (Table 3).

High UBE2C protein expression was associated with worse BCSS in the LVI-positive subgroup (p=0.048, HR =1.55, 95% CI; 1.01-2.41; Figure 3C) but not in the in the LVI-negative subgroup (p=0.526, HR= 1.81, 95% CI; 0.70-2.00; Supplementary Figure 2C).

DISCUSSION

BC is the most common malignancy affecting women worldwide [29]. LVI is a serious consequence in BC that contributes to cancer metastasis and hence shorter survival [8,9]. Despite the ability of LVI to serve as a prognostic factor in BC, the underlying mechanisms and the key molecular factors involved in BC-LVI remain unknown. UBE2C is a member of the ubiquitin-conjugating enzyme family that plays a critical role in the ubiquitin-proteasome proteolytic (UPP) pathway. Dysregulation of the UPP pathway enhances tumour oncogenes and can affect tumour suppressor proteins degradation, thereby resulting in the abnormal aggregation of those proteins in the body. Accordingly, the UPP system plays a pivotal role in cancer initiation and progression [30]. Despite the recognised importance of UBE2C in relation to cancer progression, the role played by UBE2C in BC and BC-LVI remains ill-defined.

Our study identified significant associations between high UBE2C expression and aggressive tumour characteristics, including larger tumour size, high tumour grade, lymph node positivity, NPI poor prognostic groups, LVI positivity, hormone receptor (ER and PR) negativity, high expression of the proliferative marker Ki67, p53 and HER2 positivity, and the HER2-enriched intrinsic BC subtype in addition to poor patient outcomes. These results are consistent with the results of previous studies that demonstrated that UBE2C is a key factor in cancer progression and prognosis [13, 14 , 31]. For instance, Chao-hua Mo et al investigated the prognostic significance of UBE2C expression at both transcriptomic level (1006 cases) and protein levels (209 BC tissue samples), and reported that high UBE2C expression is associated with worse outcome as well as aggressive tumour characteristics in BC [14]. High UBE2C protein expression was determined to exhibit a positive correlation with only HER2 at both the transcriptomic and proteomic levels when compared with the steroid receptors, which may suggest a correlation between UBE2C and HER2-enriched tumours when compared with the other molecular BC subtypes. The HER2-enriched BC type is considered one of the most aggressive types of BC, and it is significantly correlated with cancer cell adhesion [32,33].

The positive correlation identified in this study between UBE2C expression and the presence of both LVI and nodal status implicates UBE2C in cancer invasiveness via enhancing the EMT process, which is in accordance with a previous *in vivo* and *in vitro* study of UBE2C in non-small-cell lung cancer [31].

In endometrial cancer, the silencing of *UBE2C* plays an essential role in regulating cancer cell proliferation, migration and invasion, as well as an EMT by increasing the p53 ubiquitination and stimulating its degradation activity, thereby activating cell apoptosis and preventing carcinogenesis [7]. In addition, the dysregulation in the N-cadherin levels could stimulate MMPs production and activation to provide a proper EMT, which could lead to the enhancement of tumour cell migratory behaviour and the degradation of the primary site stroma to facilitate the cancer migration process [8]. We also found high UBE2C expression to be positively associated with a high level of EGFR, which also plays a pivotal role in BC cell migration, adhesion and invasion [34]. Moreover, high UBE2C expression might contribute to the cell adhesion process via stimulating the migration of BC tumour cells through the lymphatic vessels and starting the invasion process by activating the Wnt and PI3K signalling pathway [35]. Taken together, UBE2C could act as an essential prerequisite for BC progression that is responsible for silencing the level of E-cadherin and enhancing the levels of N-cadherin and EGFR. This may result in the activation of cancer cell migration and invasion, which may explain the vital role of UBE2C in LVI and metastasis in BC.

Moreover, the mitosis promoting factor (MPF) is an essential regulator of mitosis, which is known as an essential prerequisite for the G2/M transition [36]. In most eukaryotes, mitosis requires unique complex criteria to be activated, including formal formation, activation and cellular translocation [37,38]. Thus, an imbalance of this complex may lead to a blockage of the mitosis process and, therefore, G2/M transition. At the transcriptomic level, the high expression of *UBE2C* exhibited significant positive associations with cyclin-related genes, which play a crucial role in both the cell cycle process (G1/S and G2/M) and cell proliferation [39]. Similar to the *UBE2C* transcriptomic level results, high UBE2C protein expression showed a significant positive correlation with cyclin B1[40]. This positive correlation may indicate the critical role of UBE2C as a tumour oncogene during the cell cycle through enhancing the G1/S and G2/M transitions that prevent cancer apoptosis and promotes tumour cell proliferation via controlling the PI3K/AKT/mTOR signalling pathway [41]. In light of all this, the loss of UBE2C can lead to the blockage of the G2/M transition via downregulating the expression of CDK1 and cyclin B1 [42]. Similar results were obtained in melanoma; downregulation

of UBE2C acts as a cell growth regulator via blocking ERK/Akt signalling pathways, and preventing the G2/M transition by activating MPF and stimulating apoptosis [42]. It was also demonstrated that UBE2C plays a pivotal role in the regulation and activation of the mTOR/PI3K/AKT pathway in cervical cancer [41]. These findings support that high UBE2C expression correlates with BC progression and invasion cascades.

This study also suggests a new avenue for exploring the therapeutic role of UBE2C as an independent biomarker that could be used to target invasive BC both directly and indirectly. Targeted anti-UBC2C therapies that block UBE2C pathways would stop and/or reduce its consequent biological actions including cellular proliferation and invasiveness. In addition, therapeutic agents targeting UBE2C would synergise the effect of other therapies including chemotherapy, anti-oestrogen medications and radiation. Previous studies showed that overexpression of UBE2C reduces the therapeutic potency of letrozole, tamoxifen, doxorubicin and leads to radio-resistance in various BC cell lines [37,38,43]. These findings highlight the importance of further investigating the therapeutic and predictive potential of UBE2C expression in BC.

Although this study has presented promising findings based on evidence at both the transcriptomic and proteomic levels indicating the potentially critical role of UBE2C in BC-LVI, it is important to acknowledge that it has some limitations. First, this study was based on retrospectively collected cohort data. A well-characterised randomised clinical assessment involving more cases and uniform treatment is required for the independent evaluation of UBE2C expression in BC. Second, further *in vivo* and *in vitro* functional studies are required to discover the exact molecular mechanism(s) associated with UBE2C in order to validate its potential as a prognostic marker of BC-LVI.

In conclusion, high UBE2C expression in BC is associated with both LVI positivity and poor prognostic factors. It is an independent prognostic biomarker of poor patient survival. UBE2C may play an essential role in tumour cell proliferation, migration, invasion, and metastasis. Further *in vivo* and *in vitro* functional studies are required to investigate the molecular mechanisms of UBE2C in BC and its therapeutic potential.

CONFLICT OF INTEREST

All the authors declare that they have no conflict of interest.

ETHICAL APPROVAL

This study was approved by the Nottingham Research Ethics Committee 2 under the title 'Development of a molecular genetic classification of breast cancer' and the North West – Greater Manchester Central Research Ethics Committee under the title 'Nottingham Health Science Biobank (NHSB)' reference number 15/NW/0685. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The release of data was also pseudonymised as per the UK Human Tissue Act regulations. This article does not contain any studies with animals performed by any of the authors.

INFORMED CONSENT:

Informed consent was obtained from all individuals prior to surgery to use their tissue materials in research.

ACKNOWLEDGMENTS

The authors are part of the PathLAKE digital pathology consortium. These new Centres are supported by a £50m investment from the Data to Early Diagnosis and Precision Medicine strand of the government's Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI).

DATA AVAILABILITY

The authors confirm the data that has been used in this work is available on reasonable request.

FIGURE LEGENDS:

Figure 1: Patients' outcomes of Breast cancer survival on Transcriptomic level.

(A) Cumulative breast cancer specific survival (BCSS) of patients stratified by *UBE2C* mRNA expression in METABRIC, (B) Cumulative BCSS of patients stratified by *UBE2C* mRNA expression in TCGA, (C) Cumulative BCSS of patients stratified by *UBE2C* mRNA expression in the KM-Plotter cohort, (D) Cumulative BCSS stratified by *UBE2C* mRNA expression in LVI positive tumours in METABRIC, (E) Cumulative BCSS stratified by *UBE2C* mRNA expression in LVI positive tumours in TCGA.

Figure 2: UBE2C TMAs core protein expression. (A) UBE2C weak IHC expression. (B) UBE2C strong IHC expression in invasive breast cancer TMA cores.

Figure 3: Patients' outcomes of Breast cancer survival on UBE2C protein expression in the Nottingham cohort.

(A) Cumulative breast cancer specific survival (BCSS) of patients stratified by UBE2C protein expression. (B) Cumulative breast cancer disease free survival (BCDFS) of patients stratified by UBE2C protein expression. (C) Cumulative BCSS stratified by UBE2C protein expression in the Nottingham LVI positive cohort.

REFERENCES:

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406 (6797):747-752. doi:10.1038/35021093

2. Zhong JL, Huang CZ (2016) Ubiquitin proteasome system research in gastrointestinal cancer. World journal of gastrointestinal oncology 8 (2):198-206. doi:10.4251/wjgo.v8.i2.198

3. Hao Z, Zhang H, Cowell J (2012) Ubiquitin-conjugating enzyme UBE2C: molecular biology, role in tumorigenesis, and potential as a biomarker. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 33 (3):723-730. doi:10.1007/s13277-011-0291-1

4. Dastsooz H, Cereda M, Donna D, Oliviero S (2019) A Comprehensive Bioinformatics Analysis of UBE2C in Cancers. International journal of molecular sciences 20 (9). doi:10.3390/ijms20092228

5. Jin Z, Zhao X, Cui L, Xu X, Zhao Y, Younai F, Messadi D, Hu S (2020) UBE2C promotes the progression of head and neck squamous cell carcinoma. Biochemical and biophysical research communications 523 (2):389-397. doi:10.1016/j.bbrc.2019.12.064

6. Zhang HQ, Zhao G, Ke B, Ma G, Liu GL, Liang H, Liu LR, Hao XS (2018) Overexpression of UBE2C correlates with poor prognosis in gastric cancer patients. European review for medical and pharmacological sciences 22 (6):1665-1671. doi:10.26355/eurrev_201803_14578

7. Liu Y, Zhao R, Chi S, Zhang W, Xiao C, Zhou X, Zhao Y, Wang H (2020) UBE2C Is Upregulated by Estrogen and Promotes Epithelial-Mesenchymal Transition via p53 in Endometrial Cancer. Molecular cancer research : MCR 18 (2):204-215. doi:10.1158/1541-7786.mcr-19-0561

8. Kariri YA, Aleskandarany MA, Joseph C, Kurozumi S, Mohammed OJ, Toss MS, Green AR, Rakha EA (2020) Molecular Complexity of Lymphovascular Invasion: The Role of Cell Migration in Breast Cancer as a Prototype. Pathobiology : journal of immunopathology, molecular and cellular biology 87 (4):218-231. doi:10.1159/000508337

9. Aleskandarany MA, Sonbul SN, Mukherjee A, Rakha EA (2015) Molecular Mechanisms Underlying Lymphovascular Invasion in Invasive Breast Cancer. Pathobiology : journal of immunopathology, molecular and cellular biology 82 (3-4):113-123. doi:10.1159/000433583

10. 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):3670-3680. doi:10.1002/cncr.26711

11. Parris TZ, Kovács A, Aziz L, Hajizadeh S, Nemes S, Semaan M, Forssell-Aronsson E, Karlsson P, Helou K (2014) Additive effect of the AZGP1, PIP, S100A8 and UBE2C molecular biomarkers improves outcome prediction in breast carcinoma. International journal of cancer 134 (7):1617-1629. doi:10.1002/ijc.28497

12. Chou CP, Huang NC, Jhuang SJ, Pan HB, Peng NJ, Cheng JT, Chen CF, Chen JJ, Chang TH (2014) Ubiquitin-conjugating enzyme UBE2C is highly expressed in breast microcalcification lesions. PloS one 9 (4):e93934. doi:10.1371/journal.pone.0093934

13. Psyrri A, Kalogeras KT, Kronenwett R, Wirtz RM, Batistatou A, Bournakis E, Timotheadou E, Gogas H, Aravantinos G, Christodoulou C, Makatsoris T, Linardou H, Pectasides D, Pavlidis N, Economopoulos T, Fountzilas G (2012) Prognostic significance of UBE2C mRNA expression in high-risk early breast cancer. A Hellenic Cooperative Oncology Group (HeCOG) Study. Annals of oncology : official journal of the European Society for Medical Oncology 23 (6):1422-1427. doi:10.1093/annonc/mdr527

14. Mo CH, Gao L, Zhu XF, Wei KL, Zeng JJ, Chen G, Feng ZB (2017) The clinicopathological significance of UBE2C in breast cancer: a study based on immunohistochemistry, microarray and RNA-sequencing data. Cancer cell international 17:83. doi:10.1186/s12935-017-0455-1

15. Müller BM, Keil E, Lehmann A, Winzer KJ, Richter-Ehrenstein C, Prinzler J, Bangemann N, Reles A, Stadie S, Schoenegg W, Eucker J, Schmidt M, Lippek F, Jöhrens K, Pahl S, Sinn BV, Budczies J, Dietel M, Denkert C (2013) The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. PloS one 8 (6):e68252. doi:10.1371/journal.pone.0068252

16. Bartha Á, Győrffy B (2021) TNMplot.com: A Web Tool for the Comparison of Gene Expression in Normal, Tumor and Metastatic Tissues. International journal of molecular sciences 22 (5). doi:10.3390/ijms22052622

17. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S (2017) UALCAN: A Portal for Facilitating Tumor Subgroup Gene

Expression and Survival Analyses. Neoplasia (New York, NY) 19 (8):649-658. doi:10.1016/j.neo.2017.05.002

18. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486 (7403):346-352. doi:10.1038/nature10983

19. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C, Bowlby R, Shen H, Hayat S, Fieldhouse R, Lester SC, Tse GM, Factor RE, Collins LC, Allison KH, Chen YY, Jensen K, Johnson NB, Oesterreich S, Mills GB, Cherniack AD, Robertson G, Benz C, Sander C, Laird PW, Hoadley KA, King TA, Perou CM (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell 163 (2):506-519. doi:10.1016/j.cell.2015.09.033

20. Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast cancer research and treatment 123 (3):725-731. doi:10.1007/s10549-009-0674-9

21. Kariri YA, Joseph C, Kurozumi S, Toss MS, Alsaleem M, Raafat S, Mongan NP, Aleskandarany MA, Green AR, Rakha EA (2020) Prognostic significance of KN motif and ankyrin repeat domains 1 (KANK1) in invasive breast cancer. Breast cancer research and treatment 179 (2):349-357. doi:10.1007/s10549-019-05466-8

22. Kariri YA, Alsaleem M, Joseph C, Alsaeed S, Aljohani A, Shiino S, Mohammed OJ, Toss MS, Green AR, Rakha EA (2020) The prognostic significance of interferon-stimulated gene 15 (ISG15) in invasive breast cancer. Breast cancer research and treatment. doi:10.1007/s10549-020-05955-1

23. Muftah AA, Aleskandarany MA, Al-Kaabi MM, Sonbul SN, Diez-Rodriguez M, Nolan CC, Caldas C, Ellis IO, Rakha EA, Green AR (2017) Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections. Breast cancer research and treatment 164 (2):341-348. doi:10.1007/s10549-017-4270-0

24. Rolland P, Spendlove I, Madjd Z, Rakha EA, Patel P, Ellis IO, Durrant L (2007) The p53 positive Bcl-2 negative phenotype is an independent marker of prognosis in breast cancer. International journal of cancer 120 (6):1311-1317. doi:10.1002/ijc.22430

25. Rakha EA, Abd El Rehim D, Pinder SE, Lewis SA, Ellis IO (2005) E-cadherin expression in invasive non-lobular carcinoma of the breast and its prognostic significance. Histopathology 46 (6):685-693. doi:10.1111/j.1365-2559.2005.02156.x

26. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, 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. doi:10.1002/ijc.21004

27. McCarty KS, Jr., Miller LS, Cox EB, Konrath J, McCarty KS, Sr. (1985) Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Archives of pathology & laboratory medicine 109 (8):716-721

28. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG (2018) Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. Journal of the National Cancer Institute 110 (8):803-811. doi:10.1093/jnci/djy088

29. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A (2017) Breast cancer statistics, 2017, racial disparity in mortality by state. CA: a cancer journal for clinicians 67 (6):439-448. doi:10.3322/caac.21412

30. Doherty FJ, Dawson S, Mayer RJ (2002) The ubiquitin-proteasome pathway of intracellular proteolysis. Essays in biochemistry 38:51-63. doi:10.1042/bse0380051

31. Jin D, Guo J, Wu Y, Du J, Wang X, An J, Hu B, Kong L, Di W, Wang W (2019) UBE2C, Directly Targeted by miR-548e-5p, Increases the Cellular Growth and Invasive Abilities of Cancer Cells Interacting with the EMT Marker Protein Zinc Finger E-box Binding Homeobox 1/2 in NSCLC. Theranostics 9 (7):2036-2055. doi:10.7150/thno.32738

32. Falchook GS, Moulder SL, Wheler JJ, Jiang Y, Bastida CC, Kurzrock R (2013) Dual HER2 inhibition in combination with anti-VEGF treatment is active in heavily pretreated HER2-positive

breast cancer. Annals of oncology : official journal of the European Society for Medical Oncology 24 (12):3004-3011. doi:10.1093/annonc/mdt395

33. Schade B, Lesurf R, Sanguin-Gendreau V, Bui T, Deblois G, O'Toole SA, Millar EK, Zardawi SJ, Lopez-Knowles E, Sutherland RL, Giguère V, Kahn M, Hallett M, Muller WJ (2013) β-Catenin signaling is a critical event in ErbB2-mediated mammary tumor progression. Cancer research 73 (14):4474-4487. doi:10.1158/0008-5472.can-12-3925

34. 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):331-345. doi:10.1007/s10549-012-2289-9

35. Alsaleem M, Toss MS, Joseph C, Aleskandarany M, Kurozumi S, Alshankyty I, Ogden A, Rida PCG, Ellis IO, Aneja R, Green AR, Mongan NP, Rakha EA (2019) The molecular mechanisms underlying reduced E-cadherin expression in invasive ductal carcinoma of the breast: high throughput analysis of large cohorts. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc 32 (7):967-976. doi:10.1038/s41379-019-0209-9

36. Nurse P (1990) Universal control mechanism regulating onset of M-phase. Nature 344 (6266):503-508. doi:10.1038/344503a0

37. Porter LA, Donoghue DJ (2003) Cyclin B1 and CDK1: nuclear localization and upstream regulators. Progress in cell cycle research 5:335-347

38. Pines J (2006) Mitosis: a matter of getting rid of the right protein at the right time. Trends in cell biology 16 (1):55-63. doi:10.1016/j.tcb.2005.11.006

39. Noguchi T, Dobashi Y, Minehara H, Itoman M, Kameya T (2000) Involvement of cyclins in cell proliferation and their clinical implications in soft tissue smooth muscle tumors. The American journal of pathology 156 (6):2135-2147. doi:10.1016/s0002-9440(10)65084-5

40. Aaltonen K, Amini RM, Heikkilä P, Aittomäki K, Tamminen A, Nevanlinna H, Blomqvist C (2009) High cyclin B1 expression is associated with poor survival in breast cancer. British journal of cancer 100 (7):1055-1060. doi:10.1038/sj.bjc.6604874

41. Chiang AJ, Li CJ, Tsui KH, Chang C, Chang YI, Chen LW, Chang TH, Sheu JJ (2020) UBE2C Drives Human Cervical Cancer Progression and Is Positively Modulated by mTOR. Biomolecules 11 (1). doi:10.3390/biom11010037

42. Liu G, Zhao J, Pan B, Ma G, Liu L (2019) UBE2C overexpression in melanoma and its essential role in G2/M transition. Journal of Cancer 10 (10):2176-2184. doi:10.7150/jca.32731

43. Fuentes-Antrás J, Alcaraz-Sanabria AL, Morafraile EC, Noblejas-López MDM, Galán-Moya EM, Baliu-Pique M, López-Cade I, García-Barberán V, Pérez-Segura P, Manzano A, Pandiella A, Győrffy B, Ocaña A (2021) Mapping of Genomic Vulnerabilities in the Post-Translational Ubiquitination, SUMOylation and Neddylation Machinery in Breast Cancer. Cancers 13 (4). doi:10.3390/cancers13040833

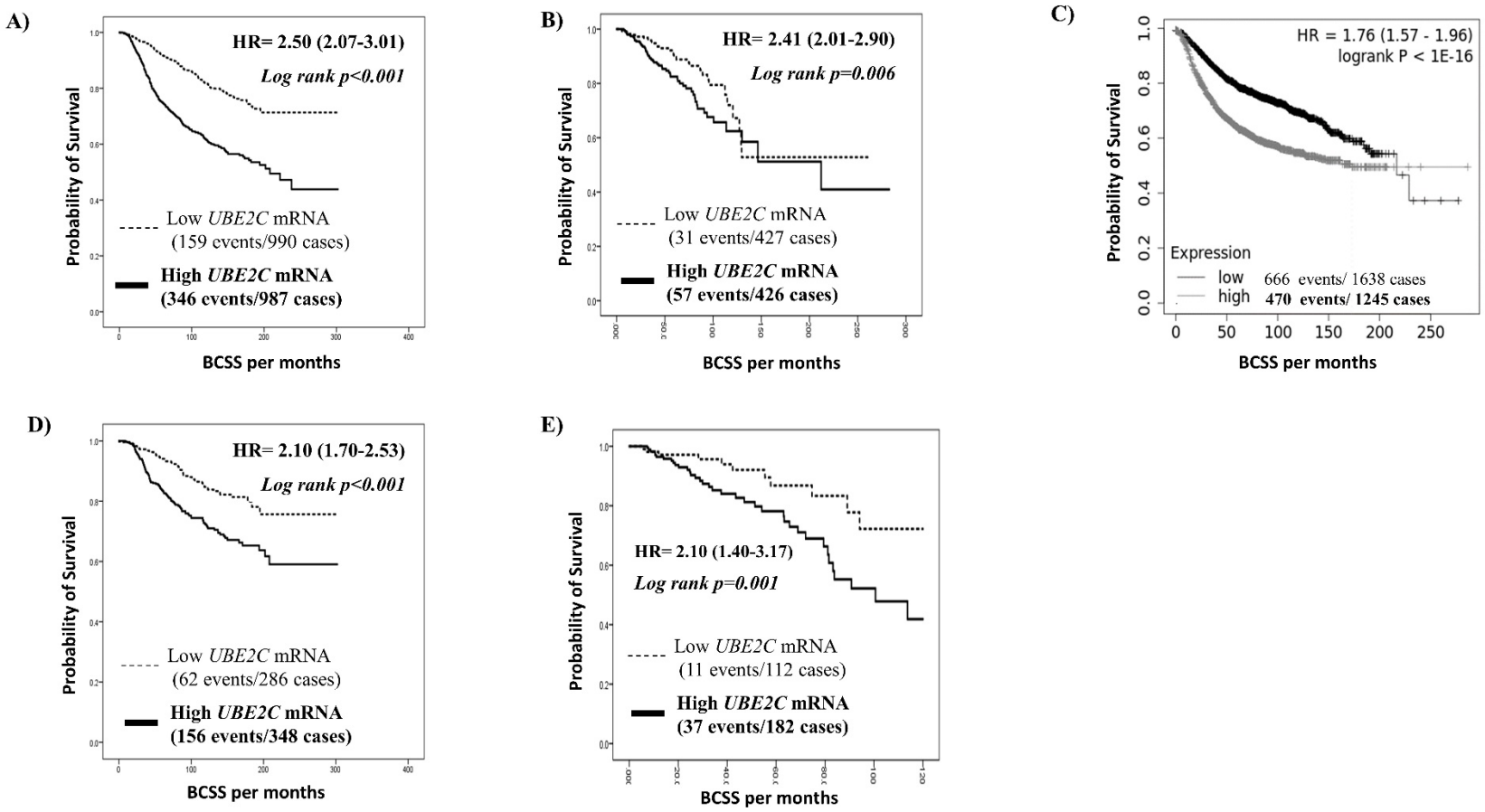


Figure 1: Patients' outcomes of Breast cancer survival on Transcriptomic level.

(A) Cumulative breast cancer specific survival (BCSS) of patients stratified by *UBE2C* mRNA expression in METABRIC, (B) Cumulative BCSS of patients stratified by *UBE2C* mRNA expression in TCGA, (C) Cumulative BCSS of patients stratified by *UBE2C* mRNA expression in the KM-Plotter cohort, (D) Cumulative BCSS stratified by *UBE2C* mRNA expression in LVI positive tumours in METABRIC, (E) Cumulative BCSS stratified by *UBE2C* mRNA expression in LVI positive tumours in TCGA.

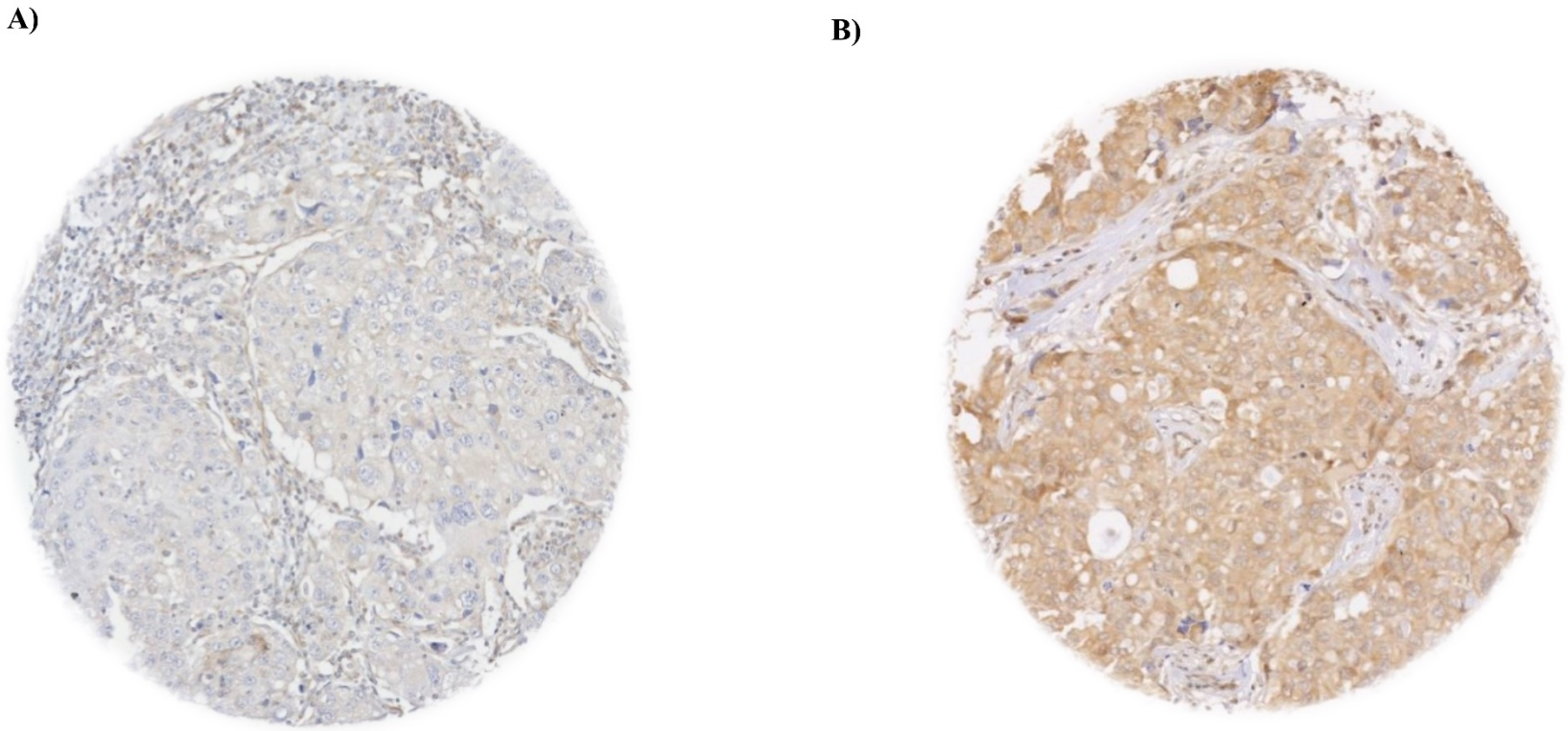


Figure 2: UBE2C TMAs core protein expression. (A) UBE2C weak IHC expression. (B) UBE2C strong IHC expression in invasive breast cancer TMA cores.

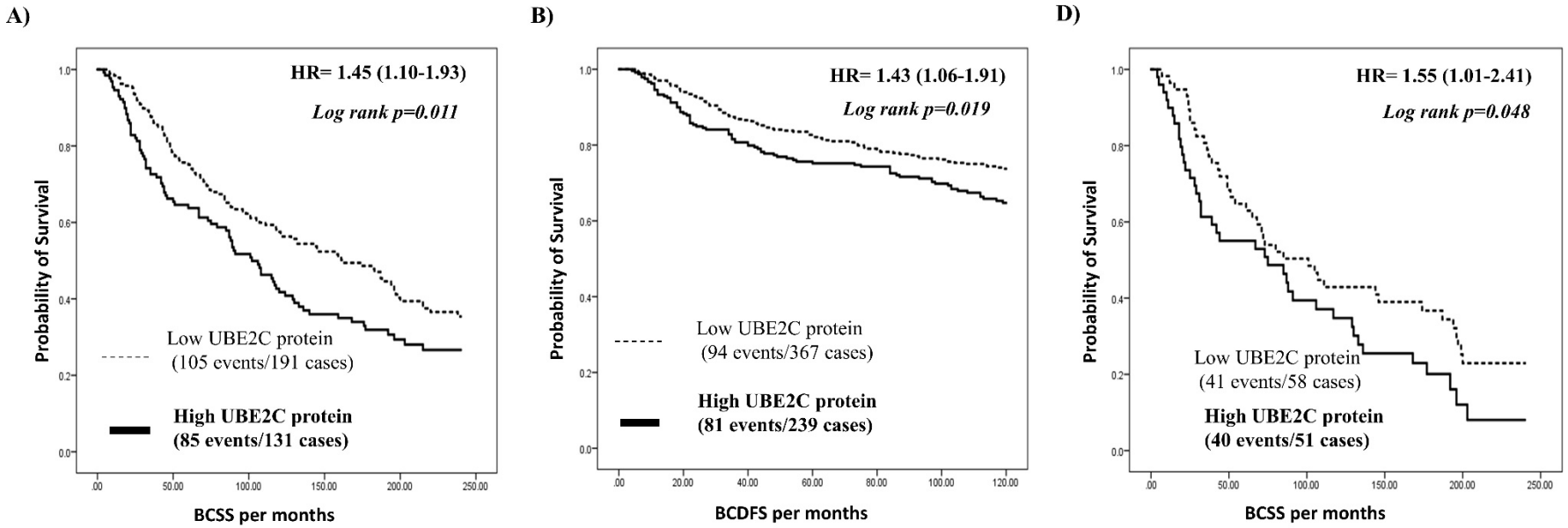


Figure 3: Patients' outcomes of Breast cancer survival on UBE2C protein expression in the Nottingham cohort.

(A) Cumulative breast cancer specific survival (BCSS) of patients stratified by UBE2C protein expression. (B) Cumulative breast cancer disease free survival (BCDFS) of patients stratified by UBE2C protein expression. (C) Cumulative BCSS stratified by UBE2C protein expression in the Nottingham LVI positive cohort.

Table 1: Association of *UBE2C* mRNA expression with clinicopathological characteristics in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (n=1980) and in the Cancer Genome Atlas (TCGA) (n=854) breast cancer series.

Parameters	METABRIC cohort			TCGA cohort		
	Low *UBE2C*	High *UBE2C*	*p* value	Low *UBE2C*	High *UBE2C*	*p* value
	N (%)	*N* (%)		*N* (%)	*N* (%)	
Tumour size						
≤ 2.0cm	492 (57)	367 (43)	**<0.001**	145 (61)	49 (39)	**<0.001**
>2.0cm	492 (45)	**609 (55)**		286 (46)	**332 (56)**	
Lymph Node status						
Negative	566 (55)	469 (45)	**<0.001**	219 (51)	207 (49)	0.471
Positive	421 (45)	**517 (55)**		207 (49)	216 (51)	
Histological grade						
Grade 1 and 2	677 (72)	263 (28)	**<0.001**	333 (72)	131 (28)	**<0.001**
Grade 3	250 (26)	**702 (74)**		71 (20)	**281 (80)**	
Tumour Histological Subtypes						
Ductal NST	684 (44)	**860 (56)**	**<0.001**	298 (51)	300 (49)	0.447
Lobular	17 (53)	15 (47)		93 (52)	84 (48)	
Medullary-like	163 (80)	40 (20)		15 (53)	13 (47)	
Special type	103 (70)	44 (30)		14 (52)	13 (48)	
Lymphovascular invasion (LVI)						
Negative	492 (53)	438 (47)	**0.002**	315 (56)	244 (44)	**<0.001**
Positive	286 (45)	**349 (55)**		113 (38)	**182 (62)**	
Oestrogen receptor (ER)						
Negative	98 (21)	**376 (79)**	**<0.001**	24 (13)	**161 (87)**	**<0.001**
Positive	892(59)	614 (41)		391 (61)	248 (39)	
Progesterone receptor (PR)						
Negative	317 (34)	**623(66)**	**<0.001**	63 (23)	**208 (77)**	**<0.001**
Positive	673 (65)	367 (35)		349 (64)	197 (36)	
Human epidermal growth factor receptor 2 (HER2)						
Negative	945 (55)	788 (45)	**<0.001**	302 (53)	265 (47)	**<0.001**
Positive	45 (18)	**202 (82)**		50 (38)	**83 (62)**	
Epithelial growth factor receptor (EFGR)						
Negative	504 (51)	486 (49)	0.418	209 (49)	218 (51)	0.494
Positive	486 (49)	504 (51)		219 (51)	208 (49)	
Molecular Subtypes						
Luminal A	614 (85)	110 (15)	**<0.001**	315 (78)	90 (22)	**<0.001**
Luminal B	130 (27)	358 (73)		23 (17)	118 (83)	
HER-enriched	45 (19)	195 (81)		9 (16)	47 (84)	
Basal-like	37 (11)	**292 (89)**		7 (5)	**126 (95)**	
Normal-like	164 (82)	35 (18)		24 (80)	6 (20)	

P values in bold means statistically significant

Table 2: Correlation of *UBE2C* mRNA expression with mRNA expression of Adhesion molecules and MMPs genes in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (n=1980) and in the Cancer Genome Atlas (TCGA) (n=854) breast cancer series.

Gene names	METABRIC cohort		TCGA cohort	
	Correlation value	*p* value	Correlation value	*p* value
Adhesion molecule genes				
CDH1	-0.093	**<0.001**	-0.020	0.553
CDH2	0.118	**<0.001**	0.046	0.179
MMPs related genes				
MMP7	0.114	**<0.001**	0.253	**<0.001**
MMP9	0.297	**<0.001**	0.152	**<0.001**
MMP12	0.303	**<0.001**	0.209	**<0.001**
MMP14	0.080	**<0.001**	0.068	**0.048**
MMP15	0.277	**<0.001**	0.190	**<0.001**
MMP20	0.137	**<0.001**	0.257	**<0.001**
MMP21	0.041	**0.040**	0.135	**<0.001**
MMP25	0.150	**<0.001**	0.119	**0.001**
Cell cycle related genes				
CDK1	0.722	**<0.001**	0.507	**<0.001**
CDK2	0.532	**<0.001**	0.392	**<0.001**
CDK4	0.400	**<0.001**	0.278	**<0.001**
CDK5	0.249	**<0.001**	0.218	**<0.001**
CDK6	0.126	**<0.001**	0.233	**<0.001**
CDKN2A	0.347	**<0.001**	0.500	**<0.001**
CCNB1	0.687	**<0.001**	0.658	**<0.001**
CCNE1	0.673	**<0.001**	0.461	**<0.001**
CCNA1	*0.220*	**<0.001**	0.116	**0.001**
CCNA2	*0.819*	**<0.001**	0.584	**<0.001**
CCNB2	*0.879*	**<0.001**	0.704	**<0.001**
CCND3	0.072	**0.001**	0.117	**0.001**
CCNE2	0.661	**<0.001**	0.323	**<0.001**
CDCA5	0.881	**<0.001**	0.580	**<0.001**
MYC	0.085	**<0.001**	0.281	**<0.001**
CDC20	0.861	**<0.001**	0.720	**<0.001**

Table 3: Multivariate Cox proportional hazard regression analysis for predictors of breast cancer specific survival (BCSS) in the METABRIC, TCGA and Nottingham BC cohort.

Factors	Breast cancer specific survival (BCSS) in METABRIC			Breast cancer specific survival (BCSS) in TCGA			Breast cancer specific survival (BCSS) in Nottingham BC cohort		
	Hazard Ratio	**95% CI**	***p* value**	**Hazard Ratio**	**95% CI**	***p* value**	**Hazard Ratio**	**95% CI**	***p* value**
UBE2C protein expression	1.9	1.50-2.38	**<0.001**	1.22	0.69-2.14	0.502	1.6	1.10 – 2.30	**0.013**
Tumour size	1.87	1.53-2.30	**<0.001**	1.24	0.68-2.30	0.483	1.34	0.93 – 5.64	0.113
Lymophvascular invasion (LVI)	1.64	1.33-2.04	**<0.001**	1.71	1.01-2.90	**0.046**	2.26	1.61 -3.17	**<0.001**
Oestrogen (ER) status	0.74	0.58-0.93	**0.009**	0.64	0.36-1.17	0.147	2.3	0.93 – 5.64	0.072
Human epidermal growth factor receptor 2 (HER2) status	1.55	1.20-2.02	**0.001**	1.32	0.71-2.47	0.384	2.6	1.61 – 4.10	**<0.001**

Significant correlations are in bold.

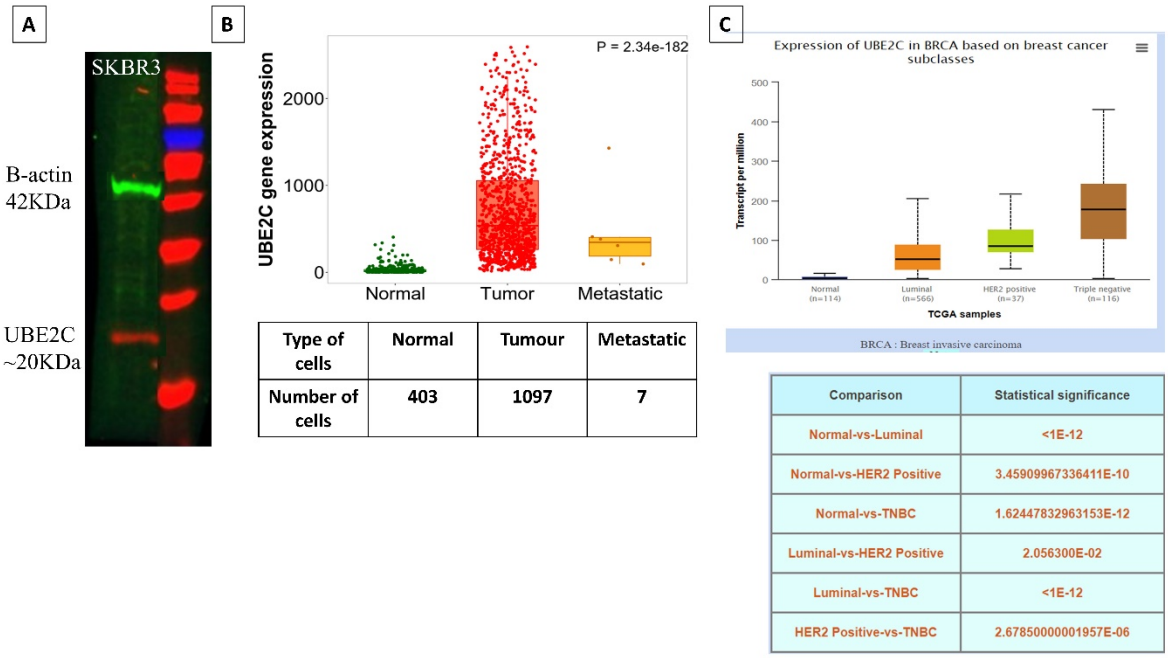
Table 4. Association between UBE2C protein expression and clinicopathological characteristics of the Nottingham breast cancer cohort (n=619).

Parameters	UBE2C protein expression		
	Low	High	*p* value
	N (%)	*N* (%)	
Tumour size			
≤ 2.0cm	192 (66)	99 (34)	**0.011**
> 2.0cm	180 (56)	**142 (44)**	
Lymph node status			
Negative	220 (62)	137 (38)	**0.019**
Positive	104 (41)	**153 (59)**	
Lymphovascular invasion (LVI)			
Negative	224 (68)	108 (32)	**0.009**
Positive	107 (56)	**84 (44)**	
Histological Grade			
Grade 1	65 (84)	12 (16)	**<0.001**
Grade 2	148 (74)	53 (26)	
Grade 3	163 (48)	**242 (52)**	
Histological Tumour Subtypes			
Ductal NST	137 (43)	**73 (33)**	**<0.001**
Lobular	100 (31)	29 (13)	
Medullary	47 (15)	67 (30)	
Special type	37 (11)	54 (24)	
Nottingham prognostic index			
Good prognostic group	124 (79)	33 (21)	**<0.001**
Moderate prognostic group	185 (54)	155 (46)	
Poor prognostic group	63 (54)	**53 (46)**	
Age			
<50	133 (63)	101 (43)	0.125
>50	239 (63)	140 (37)	
Oestrogen Receptor (ER)			
Negative	61 (38)	**99 (62)**	**<0.001**
Positive	313 (69)	143 (31)	
Progesterone Receptor (PR)			
Negative	117 (47)	**132 (53)**	**<0.001**
Positive	246 (70)	105 (30)	
Human epidermal growth factor receptor 2 (HER2)			
Negative	326 (64)	183 (36)	**<0.001**

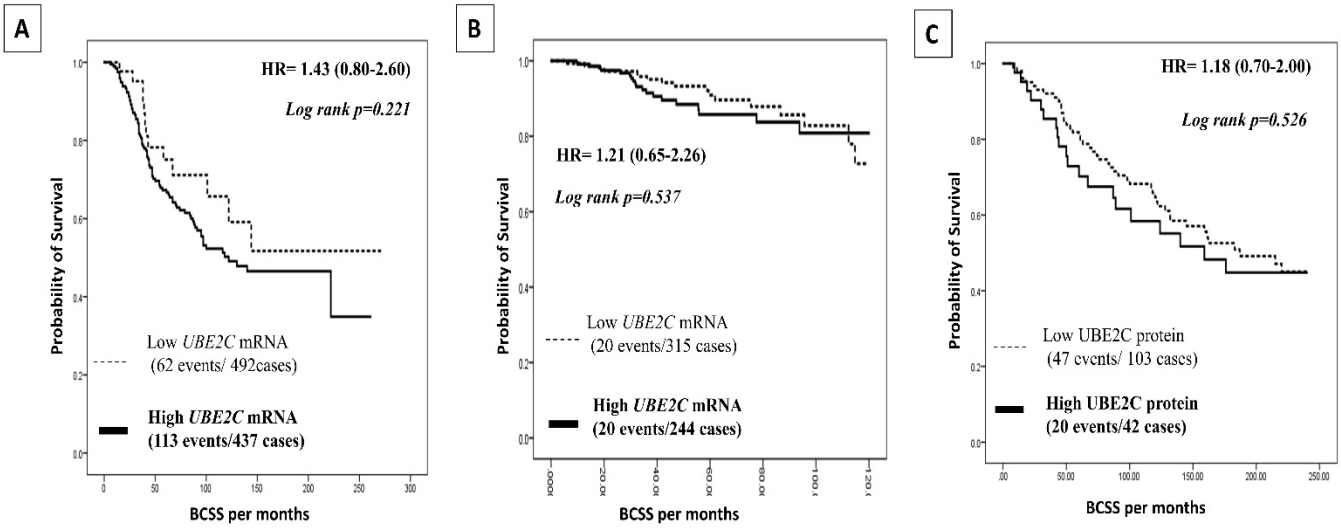
Positive	37 (41)	**54 (59)**	
P53			
Negative	283 (77)	137 (23)	**<0.001**
Positive	81 (44)	**101 (56)**	
Ki67			
Negative	139 (74)	48 (26)	**<0.001**
Positive	165 (53)	**145 (47)**	
E-Cadherin			
Negative	139 (64)	78 (36)	0.243
Positive	228 (59)	157 (41)	
N-Cadherin			
Negative	82 (66)	42 (34)	**0.033**
Positive	199 (56)	**155 (44)**	
Cyclin B1			
Negative	90 (60)	60 (40)	**0.041**
Positive	47 (41)	**67 (59)**	
Basal-phenotype			
Negative	294 (64)	167 (36)	**0.002**
Positive	68 (49)	**71 (51)**	
Epithelial growth factor receptor (EGFR)			
Negative	300 (63)	171 (36)	**0.003**
Positive	66 (49)	**68 (51)**	
CDCA5			
Negative	191 (69)	84 (31)	**0.005**
Positive	109 (45)	**135 (55)**	
PI3K			
Negative	80 (71)	33 (29)	**0.019**
Positive	217 (58)	**154 (42)**	
IHC-Subtypes			
Luminal A	137 (65)	73 (35)	**<0.001**
Luminal B	100 (77)	29 (23)	
HER2 enriched	37 (41)	**54 (59)**	
Triple negative breast cancer (TNBC)	47 (42)	**67 (58)**	

P values in bold means statistically significant.

Supplementary Figure 1:



Type of cells	Normal	Tumour	Metastatic
Number of cells	403	1097	7

Comparison	Statistical significance
Normal-vs-Luminal	<1E-12
Normal-vs-HER2 Positive	3.45909967336411E-10
Normal-vs-TNBC	1.62447832963153E-12
Luminal-vs-HER2 Positive	2.056300E-02
Luminal-vs-TNBC	<1E-12
HER2 Positive-vs-TNBC	2.67850000001957E-06

Supplementary Figure 2: