Transcriptomic and protein expression analysis reveals clinicopathological significance of Bloom's syndrome helicase (BLM) in breast cancer

Arvind Arora[1,2], Tarek MA Abdel-Fatah[1], Devika Agarwal[3], Rachel Doherty[2], Paul M Moseley[1], Mohammed A Aleskandarany[4], Andrew R Green[4], Graham Ball[3], Alaa T Alshareeda[4], Emad A Rakha[4], Stephen Y T Chan[1], Ian O Ellis[4], and Srinivasan Madhusudan[1,2*]

[1] Department of Oncology, Nottingham University Hospitals, Nottingham NG51PB, UK.

[2]Academic Unit of Oncology, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham NG51PB, UK.

[3] School of Science and Technology, Nottingham Trent University, Clifton campus, Nottingham NG11 8NS, UK.

[4] Department of Pathology, School of Medicine, University of Nottingham, Nottingham NG51PB, UK.

Running title: BLM in breast cancer

Key words: Bloom's syndrome helicase; BLM; breast cancer; prognostic factor;predictive factor

Competing financial interests: The authors declare no competing financial interests

Word count: 3750

*** Corresponding author:**

Dr Srinivasan Madhusudan

Academic Unit of Oncology

Division of Cancer and Stem Cells

School of Medicine

University of Nottingham

Nottingham University Hospitals

Nottingham NG51PB, U.K.

Telephone: +44(0)115 823 1850

Fax: +44(0)115 823 1849

E-Mail: srinivasan.madhusudan@nottingham.ac.uk

ABSTRACT

BLM has key roles in homologous recombination repair, telomere maintenance and DNA replication. Germ-line mutations in the *BLM* gene causes Bloom's syndrome, a rare disorder characterised by premature aging and predisposition to multiple cancers including breast cancer. The clinicopathological significance of BLM in sporadic breast cancers is unknown. We investigated *BLM* mRNA expression in the Molecular Taxonomy of Breast Cancer International Consortium cohort (n=1950) and validated in an external dataset of 2413 tumours. BLM protein level was evaluated in the Nottingham Tenovus series comprising 1650 breast tumours. *BLM* mRNA overexpression was significantly associated with high histological grade, larger tumour size, ER negative, PgR negative and triple negative phenotypes (ps<0.0001). *BLM* mRNA overexpression was also linked to aggressive molecular phenotypes including PAM50.Her2 (p<0.0001), PAM50.Basal (p<0.0001) and PAM50.LumB (p<0.0001) and Genufu subtype (ER+/Her2-/High proliferation) (p<0.0001). PAM50.LumA tumours and Genufu subtype (ER+/Her2-/low proliferation) were more likely to express low levels of *BLM* mRNA (ps<0.0001). Integrative molecular clusters (intClust) intClust.1 (p<0.0001), intClust.5 (p<0.0001), intClust.9 (p<0.0001) and intClust.10 (p<0.0001) were also more likely in tumours with high *BLM* mRNA expression. *BLM* mRNA overexpression was associated with poor breast cancer specific survival (BCSS) (ps<0.000001). At the protein level, altered sub-cellular localisation with high cytoplasmic BLM and low nuclear BLM was linked to aggressive phenotypes. In multivariate analysis, BLM mRNA and BLM protein levels independently influenced BCSS (p=0.03). This is the first and the largest study to provide evidence that BLM is a promising biomarker in breast cancer.

INTRODUCTION

Blooms syndrome helicase (BLM) is a key member of the RecQ family of DNA helicases and essential for the maintenance of genomic stability. BLM is an ATP-dependent 3'-5' DNA helicase involved in unwinding a variety of DNA substrates that can arise during DNA replication and repair (1-5). BLM has important roles in the initiation and regulation of homologous recombination (HR) repair of DSB (double-strand breaks). In addition, BLM is required for Holliday junction dissolution during the terminal stages of HR. To accomplish its various biological functions, BLM interacts with several DNA repair factors including topoisomerase III, hRMI1, hRMI2 and Rad51. BLM is also part of the BRCA1-associated genome surveillance complex (BASC), which contains BRCA1, MSH2, MSH6, MLH1, ATM, PMS2 and the RAD50-MRE11-NBS1 protein complex (6). In addition to its DNA repair function, BLM is involved in the processing of stalled replication forks during replication and in telomere maintenance in cells (1-5).

Bloom's syndrome (BS) is a rare disorder caused by germ-line mutation in the *BLM* gene. BS is characterised by cancer predisposition, growth retardation, immunodeficiency, sunlight hypersensitivity and impaired fertility (7). BLM germ-line mutation results in dramatic reduction in *BLM* mRNA levels and BLM protein expression leading to extensive chromosomal instability manifested classically as excessive frequency of sister chromatid exchanges (SCEs) in BS cells (1-5). BS patients are prone to develop leukemia, lymphomas and a variety of epithelial cancers including breast cancers (7). Interestingly, polymorphisms in the *BLM* gene have been associated with increased risk of development of sporadic breast cancers (8). In preclinical models, depletion of BLM by shRNA not only reduced proliferation in cells (9) but also sensitized them to chemotherapeutic agents such as

camptothecins, cisplatin, 5-fluoruracil and hydroxyurea treatment (1-5, 7). BLM is an attractive anti-cancer drug target and small molecule inhibitors of BLM are currently under pre-clinical development (10). However, target validation studies including prognostic and/or predictive significance of BLM in human sporadic tumours have not been reported and therefore remain largely unknown. We hypothesised that BLM may be dysregulated in sporadic breast cancers and influence clinical outcomes in patients. In this study, we present the first and the largest comprehensive study providing compelling evidence that altered BLM expression has prognostic and predictive significance in patients. Our data suggest that BLM is a rational target in breast cancer.

MATERIALS AND METHODS

BLM gene expression: METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohort was evaluated for *BLM* gene expression. The METABRIC study protocol, detailing the molecular profiling methodology in a cohort of 1980 breast cancer samples is described by Curtis et al (11). Patient demographics are summarized in supplementary Table S1 of supporting information. Estrogen receptor (ER) positive and/or lymphnode negative patients did not receive adjuvant chemotherapy. ER negative and/or lymphnode positive patients received adjuvant chemotherapy. RNA was extracted from fresh frozen tumours and subjected to transcriptional profiling on the Illumina HT-12 v3 platform. The data was pre-processed and normalized as described previously (11). *BLM* expression was investigated in this data set. There was only one probe for *BLM* (*BLM* probe id: ILM_1709484) in the Illumina HT-12 v3 platform. This probe has a perfect quality score as no repeat regions were targeted by the probe. The Chi-square test was used for testing association between categorical variables and a multivariate Cox model was fitted to the data using breast cancer specific death as an endpoint. Recursive partitioning was used to identify

a cut-off in gene expression values such that the resulting subgroups have significantly different survival courses.

The external validation was done using bc-GenExMiner v3.0 (Breast Cancer Gene-Expression Miner v3.0) online dataset (http://bcgenex.centregauducheau.fr) comprising previously published gene expression datasets from fifteen independent breast cancer studies totalling 2413 tumours and summarized in supplementary Table S2. The bioinformatics tool is composed of two statistical mining modules. The first module is a "prognostic module", which offers the possibility to evaluate the in vivo prognostic informativity of genes of interest in breast cancer, and the second module is a "correlation module", which permits to compute correlation coefficients between gene expressions or to find lists of correlated genes in breast cancer. We used the prognostic module in this external validation. Statistical analyses were performed by means of survival statistical tests (Cox model, Kaplan–Meier and Forest plots). Supplementary Table S2 summarizes individual cohorts where BLM mRNA expression was investigated.

BLM protein expression in breast cancer: The study was performed in a consecutive series of 1650 patients with primary invasive breast carcinomas who were diagnosed between 1986 and 1999 and entered into the Nottingham Tenovus Primary Breast Carcinoma series. Patient demographics are summarised in Supplementary Table S3. This is a well-characterized series of patients with long-term follow-up that have been investigated in a wide range of biomarker studies (12-20). All patients were treated in a uniform way in a single institution with standard surgery (mastectomy or wide local excision) with radiotherapy. Prior to 1989, patients did not receive systemic adjuvant treatment (AT). After 1989, AT was scheduled based on prognostic and predictive factor status, including Nottingham Prognostic Index (NPI), ER status, and menopausal status. Patients with NPI scores of <3.4 (low risk) did not receive AT. In pre-menopausal patients with NPI scores of ≥3.4 (high risk), classical

Cyclophosphamide, Methotrexate, and 5-Flurouracil (CMF) chemotherapy was given; patients with ER positive tumours were also offered endocrine therapy. Postmenopausal patients with NPI scores of ≥3.4 and ER positivity were offered endocrine therapy, while ER negative patients received classical CMF chemotherapy. Median follow up was 111 months (range 1 to 233 months). Overall survival data was maintained on a prospective basis. Breast cancer specific survival (BCSS) was defined as the number of months from diagnosis to the occurrence of BC related-death. Survival was censored if the patient was still alive at the time of analysis, lost to follow-up, or died from other causes. We also evaluated 20 tumour associated normal breast tissue for BLM expression.

Tumor Marker Prognostic Studies (REMARK) criteria, recommended by McShane et al (21), were followed throughout this study. Ethical approval was obtained from the Nottingham Research Ethics Committee (C202313).

Tissue Microarrays (TMAs) and immunohistochemistry (IHC): Tumours were arrayed in tissue microarrays (TMAs) constructed with 2 replicate 0.6mm cores from the centre and periphery of the tumours. The TMAs were immunohistochemically profiled for BLM and other biological antibodies (Supplementary Table S4) as previously described (12-20). Immunohistochemical staining was performed using the Thermo Scientific Shandon Sequenza chamber system (REF: 72110017), in combination with the Novolink Max Polymer Detection System (RE7280-K: 1250 tests), and the Leica Bond Primary Antibody Diluent (AR9352), each used according to the manufacturer's instructions (Leica Microsystems). The tissue slides were deparaffinised with xylene and then rehydrated through five decreasing concentrations of alcohol (100%, 90%, 70%, 50% and 30%) for two minutes each. Pre-treatment antigen retrieval was performed on the TMA sections using sodium citrate buffer (pH 6.0) and heated for 20 minutes at 95°C in a microwave (Whirpool JT359 Jet Chef

1000W). A set of slides were incubated for 18 hours with the primary anti-BLM antibody (NBP1-89929, Novus Biologicals, UK), at a dilution of 1:100. Negative and positive (by omission of the primary antibody and IgG-matched serum) controls were included in each run. The negative control ensured that all the staining was produced from the specific interaction between antibody and antigen.

Evaluation of immune staining: The tumour cores were evaluated by two scorers (TAF and AA) and the concordance between the two scorer was excellent (k = 0.79). Whole field inspection of the core was scored and intensities of nuclear staining were grouped as follows: 0 = no staining, 1 = weak staining, 2 = moderate staining, 3 = strong staining. The percentage of each category was estimated (0-100%). Histochemical score (H-score) (range 0-300) was calculated by multiplying intensity of staining and percentage staining. A median H score of ≥ 50 was taken as the cut-off for high BLM nuclear and cytoplasm expression. Not all cores within the TMA were suitable for IHC analysis as some cores were missing or lacked tumour (<15% tumour).

Statistical analysis: Data analysis was performed using SPSS (SPSS, version 17 Chicago, IL). Where appropriate, Pearson's Chi-square, Fisher's exact, Student's t and ANOVA one way tests were used. Cumulative survival probabilities were estimated using the Kaplan–Meier method, and differences between survival rates were tested for significance using the log-rank test. Multivariate analysis for survival was performed using the Cox proportional hazard model. The proportional hazards assumption was tested using standard log-log plots. Hazard ratios (HR) and 95% confidence intervals (95% CI) were estimated for each variable. All tests were two-sided with a 95% CI and a p value < 0.05 considered significant. For multiple comparisons, p values were adjusted according to Benjamini-Hochberg method (22).

Breast cancer cell lines and culture: MCF-7 (ER+/PR+/HER2-, BRCA1 proficient), MDA-MB-231 (ER-/PR-/HER2-, BRCA1 proficient), MDA-MB-468 (ER-/PR-/HER2-, BRCA1 proficient) and MDA-MB-436 (ER-/PR-/HER2-, BRCA1 deficient) were used in the current study. All cell lines were purchased from ATCC and authenticated by ATCC. Cells were grown in RPMI (MCF-7, MDA-MB-231) or DMEM (MDA-MB-468 and MDA-MB-436) medium with the addition of 10% foetal bovine serum and 1% penicillin/streptomycin. Cell lysates were prepared and Western blot analysis performed. Primary anti-BLM antibody (NBP1-89929, Novus Biologicals, and UK) was incubated over night at room temperature at a dilution of 1:1500. Primary anti-β actin antibody (1:10000 dilution [Abcam]) was used as a loading control. Infrared dye-labelled secondary antibodies (Li-Cor) [IRDye 800CW Mouse Anti-Rabbit IgG and IRDye 680CW Rabbit Anti-Mouse IgG] were incubated at a dilution of 1:10000 for 1 hour. Membranes were scanned with a Li-Cor Odyssey machine (700 and 800nm) to determine protein expression.

Quantitative real –time PCR: Total RNA was extracted from MCF-7, MDA-MB-231, MDA-MB-468 and MDA-MB-436 cells using RNeasy Mini kit (QIAGEN, UK). The quantification of the extracted RNA was done using a NanoDrop 2000c Spectrophotometer (Thermo Scientific, UK). The cDNA was synthesized from 0.5 μg of total RNA using RT^2 first strand kit (QIAGEN, UK). qPCR was performed using SYBR Green PCR Master mix (applied biosystems,Warrington,UK) with primer set (BLM QuantiTect Prier Assay,Cat. No. QT00027671, QIAGEN) targeting BLM gene. The glyceraldehyde-3-phosphate dehydrogenase housekeeper gene was used as an internal control (GAPDH QuantiTect Prier Assay, Cat. No. QT00079247, QIAGEN). The real-time PCR for each RNA sample was performed in triplicate. NTC (No Template Control) was used to rule out cross contamination

of reagents and surfaces. NTC included all the RT-PCR reagents except the RNA template. Minus reverse transcriptase (- RT) control was used to rule out genomic DNA contamination.

RESULTS

High *BLM* transcript levels correlate to aggressive breast cancer

BLM mRNA level was investigated in the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohort comprising 1980 breast tumours. High *BLM* mRNA expression was highly significantly associated with aggressive clinicopathological features (Table 1) including high histological grade, larger tumour size, high-risk Nottingham prognostic index (NPI >3.4), Her-2 over expression, ER negative, PR negative and triple negative phenotypes (ps<0.0001). High *BLM* mRNA expression was also found to be significantly associated with previously described molecular phenotypes in breast cancer: PAM50.Her2 (p<0.0001), PAM50.Basal (p<0.0001) and PAM50.LumB (p<0.0001), Genufu subtype (ER-/Her2-), Genufu subtype (ER+/Her2-/High proliferation) and Genufu subtype (Her2 positive) breast tumours. However, PAM50.LumA tumours and Genufu subtype (ER+/Her2-/low proliferation) were more likely to express low levels of *BLM* mRNA (ps<0.0001). Similarly, *BLM* mRNA level was significantly associated with the various biological subgroups [labelled integrative clusters (intClust) 1-10] described in the METABRIC study which was based on gene copy number changes and gene expression data (11). High *BLM* mRNA expression was significantly associated with intClust.1 (p<0.0001), intClust.5 (p<0.0001), intClust.9 (p<0.0001) and intClust.10 (p<0.0001), which had the worst clinical outcome in the METABRIC study (11). Low *BLM* mRNA expression was associated

with intClust.3 ($p<0.0001$), intClust.4 ($p<0.0001$), intClust.7 ($p=0.003$) and intClust.8 ($p<0.0001$), which had intermediate to good prognosis in the METABRIC study (11).

We then proceeded to survival analysis. High *BLM* mRNA expression in tumours was associated with adverse BCSS in the whole cohort ($p<0.0001$) (Figure 1A). In ER+ sub-group, high *BLM* mRNA expression was associated with poor BCSS ($p<0.0001$) (Figure 1B). In the ER+ sub-group that received adjuvant endocrine therapy, high *BLM* mRNA expression remains associated with poor BCSS ($p<0.0001$) (Figure 1D). In ER- sub-group, low *BLM* mRNA expression was associated with poor BCSS with borderline significance ($p=0.049$) (Figure 1C). In the ER- sub-group that received adjuvant chemotherapy, although there was a trend, *BLM* mRNA expression did not significantly influence outcome ($p=0.062$) (Figure 1E) and was most likely due to limited number of patients in this cohort (n=262). In multivariate Cox regression analysis that included other validated prognostic factors, such as lymph node stage, histological grade and tumour size , *BLM* mRNA expression was a powerful independent predictor for breast cancer specific survival ($p<0.00001$) (Table 2). External validation was performed using bc-GenExMiner v3.0 (Breast Cancer Gene-Expression Miner v3.0) online dataset (http://bcgenex.centregauducheau.fr) comprising previously published gene expression datasets from fifteen independent breast cancer studies totalling 2413 tumours and summarized in supplementary materials and Table S2. The dataset provides information on metastasis relapse (MR) free survival data. As shown in the Forest plot (Supplementary Figure S1) low *BLM* mRNA expression was significantly associated with better MR free survival (Supplementary Figure S1A and S1B). Taken together, the data provides the first compelling evidence that high BLM *mRNA* expression has prognostic and/or predictive significance in breast cancer.

Sub-cellular localisation of BLM protein is associated with aggressive breast cancer

BLM is a 1417 amino acid protein with a highly conserved centrally located helicase domain. In addition, BLM has multiple domains involved in DNA- binding, ATPase activity and interaction with other binding partners. The nuclear localisation signal is present in the C-terminal region of the protein (1-5). BLM is primarily expressed in late S/G2 phase of the cell cycle. Upon DNA damage BLM localises to the nucleus where it interacts with Rad51 and is intimately involved in HR repair that is operational during the S-phase of the cell cycle (23). In addition, BLM undergoes post translational modifications such as phosphorylation and SUMOylation that can affect intracellular localisation and biochemical activity (1-5). Besides a role in HR repair, BLM is also known to interact with key factors involved in base excision repair (BER) (e.g. FEN1) and non-homologous end joining pathway (NHEJ) (e.g. DNA-PKcs) (1-5). Moreover, BLM also interacts with important players in DNA- damage signalling and cell cycle regulation (ATM-Chk2 and ATR-Chk1 pathway), which ultimately dictate whether a cell initiates cell cycle arrest to allow DNA repair or proceed to apoptosis (1-5). We therefore investigated BLM protein expression in breast cancer and correlated to expression of other markers associated the DNA-damage signalling, NHEJ, BER, cell cycle regulation and apoptosis.

We proceeded to evaluation of BLM protein expression in breast cancers. We initially profiled a panel of breast cancer cell lines. As shown in Supplementary Figure S2A; MDA-MB-231, MDA-MB-436 and MDA-MB-468 breast cancer cells have robust expression of BLM protein. In contrast, MCF-7 has low BLM expression. At the mRNA level, MCF-7 cells have low *BLM* mRNA compared MDA-MB-231, MDA-MB-436 and MDA-MB-468 cells. The data demonstrates differential BLM expression across different breast cancer cell lines.

We then conducted immunohistochemical evaluation of BLM protein expression in the Nottingham Tenovus series comprising 1650 breast tumours. Surprisingly, we observed complex sub-cellular localization of BLM protein in breast cancers including tumours exhibiting nuclear staining only, cytoplasmic staining only, nuclear-cytoplasmic co-expression or negative staining. We also evaluated 20 tumour associated normal breast tissue for BLM expression. We observed strong nuclear staining in 19/20 normal breast tissue (mean H-score =235) (supplementary Figure S2B1). 1/20 did not show any nuclear BLM staining. No cytoplasmic staining was observed in any normal breast tissue. The data confirms that nuclear expression is a common feature of normal breast tissue and altered sub-cellular localisation is a feature of breast tumours.

Nuclear BLM protein level and breast cancer: Low nuclear BLM levels were seen in 54% of tumours (n= 682/1253) and high nuclear BLM levels were observed in 46% of tumours (n= 571/1253) (Supplementary Figure S2B4). As shown in supplementary table S5, low nuclear BLM level was significantly associated with larger tumours, high tumour grade, higher mitotic index, pleomorphism and tumour type ($p<0.05$). ER-, PR-, AR-, triple negative and basal-like phenotypes were more common in tumours with low nuclear BLM protein level ($p<0.01$). BRCA1 negative, low XRCC1, low FEN1, low SMUG1, low APE1, low Polβ, low ATR and low DNA-PKcs were significantly associated with tumours that have low nuclear BLM protein level. In addition, high p16, low p21, high MIB1, high p53, low Bcl-2, low Top2A, low nuclear pCHEK1 and low nuclear Chk2 were more common in tumours with low nuclear BLM protein level ($p<0.05$).

Cytoplasmic BLM protein level and breast cancer: High cytoplasmic BLM levels were seen in 53% of tumours (n= 642/1212) and low cytoplasmic BLM levels were seen in 47% of

tumours (n= 570/1212) (Supplementary Figure S2B3). As shown in supplementary table S6, high cytoplasmic BLM level was significantly associated with pleomorphism, tumour type, high XRCC1, high FEN1, high APE1, high ATR, high DNA-PKcs, high MIB1, high Chk2, high Bax levels.

Nuclear and cytoplasmic co-expression of BLM in breast cancer: 28% (333/1253) of tumours were low nuclear/high cytoplasmic, 26.5% (332/1253) were low nuclear/low cytoplasmic, 26.5% (333/1253) were high nuclear/high cytoplasmic and 19% (238/1253) were high nuclear/low cytoplasmic (Supplementary Figure S2B5). Clinicopathological association are shown in Table 3 and supplementary Table S7. Tumours with high cytoplasmic/low nuclear BLM levels were more likely to be high grade, high mitotic index, pleomorphism, IDC-NST tumour type, PR-, triple negative and basal-like phenotype tumours ($p<0.0001$). High p16, low p21, high MIB1, high p53 and high Bax levels more common in tumours with high cytoplasmic/low nuclear BLM levels. We also correlated BLM co-expression with various DNA repair factors and observed significant associations. BRCA1 negativity was observed in 24.6% of BLM n-/c- tumours compared to 13.2% (BLM n+/c- tumours), 20.3% (BLM n-/c+ tumours) and 17.3% (BLM n+/c+ tumours). Similarly, BLM n-/c- tumours were more likely to exhibit low XRCC1 (25.6%), low FEN1 (83.8%), low SMUG1 (47.1%), low APE1 (66.8%), low pol β (50.9%), low ATR (75.9%) and DNA-PKcs (45.8%) compared to tumours that express BLM n+/c-, BLM n-/c+, or BLM n+/c+ co-expression (see Table 3).

BLM and Rad51 protein co-expression in breast cancer: A key interacting partner of BLM is Rad51 (24). Together BLM-Rad51 play an essential role in HR repair (1-5). We therefore conducted exploratory nuclear co-expression studies in breast cancer. As shown in

supplementary Table S8, we observed significant association between BLM-/Rad51- tumours and NPI>3.4, high grade, high mitotic index, pleomorphism, tumour type. Interestingly, ER negativity was observed in 47.1% of BLM-/RAD51- tumours compared to 30.5% (BLM+/RAD51- tumours), 30.9% (BLM-/RAD51+ tumours) and 17.9% in BLM+/RAD51+ tumours. Similarly, PR negativity was observed in 64.4% of BLM -/RAD51- tumours compared to 47.9% (BLM+/RAD51- tumours), 42.9% (BLM-/RAD51+ tumours) and 35.3% (BLM+/RAD51+ tumours) (see supplementary Table S8).

Survival analyses: In univariate analysis, in high risk ER positive tumours that received no endocrine therapy, patients whose tumours had high nuclear/low cytoplasmic BLM had poor BCSS (p=0.036) implying that altered expression has prognostic significance (Supplementary Figure S3). In patients who received endocrine therapy, although low nuclear/high cytoplasmic BLM tumours have the worst survival status in breast cancer, there was no statistical significance. Similarly in ER- tumours, BLM level did not significantly influence survival. When BLM (nuclear) and Rad51 (nuclear) were investigated together, BLM-/Rad51- tumours have poor survival in the whole cohort and in the ER- sub-group that received adjuvant chemotherapy (Supplementary Figure S4). BLM/Rad51 expression did not influence survival in ER + tumours (Supplementary Figure S5). In multivariate analysis (Supplementary Table S9), nuclear BLM level independently influenced survival (p=0.026). Tumour stage, grade and HER-2 expression were other factors independently associated with breast cancer specific survival.

DISCUSSION

DNA helicases are molecular motors that unwind DNA, a process that is required during DNA replication, DNA repair and telomere maintenance. RecQ family of DNA helicases

includes RECQL1, RECQL4, RECQL5, WRN and BLM. The critical role played by RecQ family of DNA helicases in genomic stability is underpinned by the fact that germ-line mutations in these genes result in genetic disorders characterised by premature aging and/or predisposition to cancers (1-5). RecQ helicases may also have a role in the pathogenesis of sporadic cancers. RECQL4 has been shown to be involved in prostate carcinogenesis (1-5). *RECQL1* genetic polymorphisms have been linked to pancreatic cancer and RECQL1 overexpression has been demonstrated in head & neck and brain tumours (1-5). In the current study, we have comprehensively investigated the role of BLM in breast cancer. We provide compelling evidence that high *BLM* mRNA expression is a strong prognostic and predictive biomarker in breast cancer. High *BLM* mRNA was linked to aggressive clinicopathological phenotypes. High *BLM* mRNA was associated with aggressive molecular phenotypes including PAM50. Luminal B, PAM50. Her2 and PAM50. basal molecular phenotypes. Given the role of BLM during replication and proliferation (25), it is perhaps not surprising that high *BLM* mRNA was more frequent in aggressive breast cancers. To further support this hypothesis we also observed that low *BLM* mRNA expression was more common in PAM50. Lumina A and ER+/Her-2 negative/low proliferation Genefu subtype tumours. Interestingly, *BLM* mRNA levels are also linked to biologically distinct integrative clusters reported in the METABRIC study (11). High *BLM* mRNA level was frequent in intClust 10 subgroup which is the most highly genomically unstable sub group with basal-like features. Low *BLM* mRNA level was seen in intClust 3 subgroup that is characterised by low genomic instability. Together the data suggest that *BLM* mRNA level may also inform genomic stability status in breast. In addition, high *BLM* mRNA level is also frequently seen in intClust 5 (HER-2 enriched with worst survival), intClust 9 (8q cis-acting/20qamplified mixed subgroup), and intClust 1 (17q23/20q cis-acting luminal B subgroup) subgroups that also manifest an aggressive phenotype. On the other hand, low *BLM* mRNA level is linked to intClust 4

(includes both ER-positive and ER-negative cases with a flat copy number landscape and termed the 'CNA-devoid' subgroup with extensive lymphocytic infiltration), intClust 7 (16p gain/16q loss with higher frequencies of 8q amplification luminal A subgroup) and intClust 8 subgroups (classical 1q gain/16q loss luminal A subgroup) (11). Of note, the data presented here is strikingly similar to the clinicopathological associations we recently reported for *FEN1* (flap endonuclease 1), a key player in long-patch base excision repair and DNA replication, in the METABRIC cohort (15). Interestingly, BLM has been shown to stimulate FEN1 activity in a preclinical study (26). The functional interaction appeared to be independent of BLM helicase activity in that study (26).

At the protein level, low nuclear and/or high cytoplasmic expression was associated with aggressive phenotypes. Association with high cytoplasmic expression was surprising. In contrast, normal breast tissue showed only strong nuclear staining and no cytoplasmic staining. As cytoplasmic function of BLM has not been described previously, we speculate that cytoplasmic accumulation in a proportion of breast tumours probably reflects dysregulation of mechanisms involved in nuclear localization of BLM. Cytoplasmic accumulation along with low nuclear BLM expression could then increase genomic instability in tumours and promote a mutator phenotype characterised by aggressive biology. To support this hypothesis we observed that low nuclear BLM levels with or without cytoplasmic expression were more likely to be high grade, high mitotic index, pleomorphism, IDC-NST tumour type, PR-, triple negative and basal-like phenotype tumours. In addition, low nuclear BLM with or without cytoplasmic expression was also associated with impaired expression of other DNA repair factors including BRCA1 negativity, low XRCC1, low FEN1, low SMUG1, low APE1, low Polβ, low ATR and low DNA-PKcs. Moreover, in multivariate analysis, nuclear BLM level independently influenced survival. As BLM and

Rad51 are known to interact with each other for efficient HR repair (24), we also performed BLM-Rad51 co-expression studies. As expected, low nuclear BLM/low nuclear RAD51 tumours exhibited aggressive phenotype and associated with poor survival. In a previous small study in normal and neoplastic human cells, BLM protein expression was shown to be overexpressed in a panel of tumour tissue compared to normal tissue including a cohort of nine breast tumours (27). Similar to our study, the authors observed a positive correlation between BLM and Ki67 but did not report any clinicopathological associations (27). Another interesting observation in the current study was that although *BLM* mRNA overexpression was categorically associated with aggressive tumours and poor outcomes, at the protein level, the association appeared more complex with low nuclear BLM protein level or low nuclear/high cytoplasmic BLM protein level being associated with adverse features. We speculate that either BLM mRNA is subjected to post-transcriptional regulation or post translational dysregulation of BLM protein expression/sub-cellular localization could in turn affect *BLM* mRNA expression through feedback loops. Detailed mechanistic studies are therefore required to understand the regulation of BLM in vivo. Data presented in the current study also suggest that BLM could be a promising marker for personalization of therapy. As low BLM is a marker of impaired HR repair, we would argue that low BLM tumours could be targeted by synthetic lethality using inhibitors of base excision repair such as those targeting PARP (28). Alternatively high BLM tumours could be targeted by small molecular inhibitors of BLM that are currently under development (10). In conclusion we provide the first clinical evidence that BLM is a promising biomarker and a rational drug target in breast cancer.

REFERENCES

1. Sharma S, Doherty KM,Brosh RM, Jr. Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. Biochem J. 2006;398:319-37.
2. Payne M,Hickson ID. Genomic instability and cancer: lessons from analysis of Bloom's syndrome. Biochem Soc Trans. 2009;37:553-9.
3. Croteau DL, Popuri V, Opresko PL,Bohr VA. Human RecQ helicases in DNA repair, recombination, and replication. Annu Rev Biochem. 2014;83:519-52.
4. Chu WK,Hickson ID. RecQ helicases: multifunctional genome caretakers. Nat Rev Cancer. 2009;9:644-54.
5. Brosh RM, Jr. DNA helicases involved in DNA repair and their roles in cancer. Nat Rev Cancer. 2013;13:542-58.
6. Wang Y, Cortez D, Yazdi P, Neff N, Elledge SJ,Qin J. BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev. 2000;14:927-39.
7. Cheok CF, Bachrati CZ, Chan KL, Ralf C, Wu L,Hickson ID. Roles of the Bloom's syndrome helicase in the maintenance of genome stability. Biochem Soc Trans. 2005;33:1456-9.
8. Sassi A, Popielarski M, Synowiec E, Morawiec Z,Wozniak K. BLM and RAD51 genes polymorphism and susceptibility to breast cancer. Pathol Oncol Res. 2013;19:451-9.
9. Mao FJ, Sidorova JM, Lauper JM, Emond MJ,Monnat RJ. The human WRN and BLM RecQ helicases differentially regulate cell proliferation and survival after chemotherapeutic DNA damage. Cancer Res. 2010;70:6548-55.
10. Nguyen GH, Dexheimer TS, Rosenthal AS, Chu WK, Singh DK, Mosedale G, et al. A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. Chem Biol. 2013;20:55-62.
11. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486:346-52.
12. Albarakati N, Abdel-Fatah TM, Doherty R, Russell R, Agarwal D, Moseley P, et al. Targeting BRCA1-BER deficient breast cancer by ATM or DNA-PKcs blockade either alone or in combination with cisplatin for personalized therapy. Mol Oncol. 2015;9:204-17.
13. Abdel-Fatah TM, Perry C, Arora A, Thompson N, Doherty R, Moseley PM, et al. Is There a Role for Base Excision Repair in Estrogen/Estrogen Receptor-Driven Breast Cancers? Antioxid Redox Signal. 2014;21:2262-8.
14. Abdel-Fatah T, Arora A, Agarwal D, Moseley P, Perry C, Thompson N, et al. Adverse prognostic and predictive significance of low DNA-dependent protein kinase catalytic subunit (DNA-PKcs) expression in early-stage breast cancers. Breast Cancer Res Treat. 2014;146:309-20.
15. Abdel-Fatah TM, Russell R, Albarakati N, Maloney DJ, Dorjsuren D, Rueda OM, et al. Genomic and protein expression analysis reveals flap endonuclease 1 (FEN1) as a key biomarker in breast and ovarian cancer. Mol Oncol. 2014;8:1326-38.
16. Abdel-Fatah TM, Russell R, Agarwal D, Moseley P, Abayomi MA, Perry C, et al. DNA polymerase beta deficiency is linked to aggressive breast cancer: a comprehensive analysis of gene copy number, mRNA and protein expression in multiple cohorts. Mol Oncol. 2014;8:520-32.
17. Abdel-Fatah TM, Perry C, Moseley P, Johnson K, Arora A, Chan S, et al. Clinicopathological significance of human apurinic/apyrimidinic endonuclease 1 (APE1) expression in oestrogen-receptor-positive breast cancer. Breast Cancer Res Treat. 2014;143:411-21.

18. Abdel-Fatah TM, Albarakati N, Bowell L, Agarwal D, Moseley P, Hawkes C, et al. Single-strand selective monofunctional uracil-DNA glycosylase (SMUG1) deficiency is linked to aggressive breast cancer and predicts response to adjuvant therapy. Breast Cancer Res Treat. 2013;142:515-27.
19. Abdel-Fatah TM, Perry C, Dickinson P, Ball G, Moseley P, Madhusudan S, et al. Bcl2 is an independent prognostic marker of triple negative breast cancer (TNBC) and predicts response to anthracycline combination (ATC) chemotherapy (CT) in adjuvant and neoadjuvant settings. Ann Oncol. 2013;24:2801-7.
20. Sultana R, Abdel-Fatah T, Abbotts R, Hawkes C, Albarakati N, Seedhouse C, et al. Targeting XRCC1 deficiency in breast cancer for personalized therapy. Cancer Res. 2013;73:1621-34.
21. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M,Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst. 2005;97:1180-4.
22. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6:65-70.
23. Bischof O, Kim SH, Irving J, Beresten S, Ellis NA,Campisi J. Regulation and localization of the Bloom syndrome protein in response to DNA damage. J Cell Biol. 2001;153:367-80.
24. Braybrooke JP, Li JL, Wu L, Caple F, Benson FE,Hickson ID. Functional interaction between the Bloom's syndrome helicase and the RAD51 paralog, RAD51L3 (RAD51D). J Biol Chem. 2003;278:48357-66.
25. Kawabe T, Tsuyama N, Kitao S, Nishikawa K, Shimamoto A, Shiratori M, et al. Differential regulation of human RecQ family helicases in cell transformation and cell cycle. Oncogene. 2000;19:4764-72.
26. Sharma S, Sommers JA, Wu L, Bohr VA, Hickson ID,Brosh RM, Jr. Stimulation of flap endonuclease-1 by the Bloom's syndrome protein. J Biol Chem. 2004;279:9847-56.
27. Turley H, Wu L, Canamero M, Gatter KC,Hickson ID. The distribution and expression of the Bloom's syndrome gene product in normal and neoplastic human cells. Br J Cancer. 2001;85:261-5.
28. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005;434:913-7.

Table 1: Association between *BLM* mRNA expression and clinico-pathologic variables in METABRIC cohort.

Variable	BLM mRNA Expression		P Values	
	Low (n=763)	High (n=1208)	*Unadjusted*	*Adjusted**
	N (%)	N (%)		
A) Pathological Parameters				
Lymph node stage				
Negative	434(56.9%)	601(49.8%)	**0.003**	**0.0034**
Positive (1-3)	100(13.1%)	214(17.7%)		
Positive (>3)	229(30.0%)	393(32.5%)		
Grade				
G1	124(17.3%)	45(3.8%)	**$1.9X10^{-63}$**	**$1.0X10^{-5}$**
G2	404(56.3%)	366(31.3%)		
G3	190(26.5%)	760(64.9%)		
Tumour Size (cm)				
T 1a+b(1.0)	49(6.4%)	43(3.6%)	**$1.4X10^{-5}$**	**$1.0X10^{-5}$**
T 1c(>1.0-2.0)	334(43.9%)	432(36.1%)		
T2 (>2.0-5)	341(44.9%)	660(55.1%)		
T3 (>5)	36(4.7%)	62(5.2%)		
NPI				
≤ 3.4	385(50.3%)	295(24.3%)	**$2.2X10^{-32}$**	**$1.0X10^{-5}$**
>3.4	380(49.7%)	917(75.7%)		
Her2 overexpression (No)	733(95.8%)	999(82.4%)	**$1.3X10^{-18}$**	**$1.0X10^{-5}$**
(Yes)	32(4.2%)	213(17.6%)		
Triple negative (No)	731(95.6)	929 (76.7)	**$6.5X10^{-29}$**	**$1.0X10^{-5}$**
(Yes)	34(4.4)	283(23.3)		
ER (Negative)	55(7.2%)	415(34.2%)	**$4.3X10^{-43}$**	**$1.0X10^{-5}$**
(Positive)	710(92.8%)	797(65.8%)		
PR (Negative)	223(29.2%)	713(58.8%)	**$6.4X10^{-38}$**	**$1.0X10^{-5}$**
(Positive)	542(70.8%)	499(41.2%)		
Genefu subtype				
ER-/Her-2 negative	20(5.1%)	130(21.5%)	**$2.2X10^{-12}$**	**$1.0X10^{-5}$**
ER+/Her-2 negative/high proliferation	71(18.3%)	295(48.8%)	**$2.2X10^{-22}$**	**$1.0X10^{-5}$**
ER+/Her-2 negative/low proliferation	283(72.8%)	85(14.0%)	**$4.4X10^{-78}$**	**$1.0X10^{-5}$**
Her-2 positive	15(3.9%)	95(15.7%)	**$6.2X10^{-9}$**	**$1.0X10^{-5}$**
PAM50 subtype				
PAM50.Her2	33(5.2%)	205(18.0%)	**$3.8X10^{-14}$**	**$1.0X10^{-5}$**
PAM50.Basal	19(3.0%)	311(27.3%)	**$2.2X10^{-36}$**	**$1.0X10^{-5}$**
PAM50.LumA	483(76.2%)	232(20.4%)	**$8.1X10^{-117}$**	**$1.0X10^{-5}$**
PAM50.LumB	98(15.5%)	391(34.3%)	**$1.7X10^{-17}$**	**$1.0X10^{-5}$**
IntClust subgroups				
intClust.1	21(2.7%)	116(9.6%)	**$5.8X10^{-9}$**	**$1.0X10^{-5}$**
intClust.2	20(2.6%)	52(4.3%)	0.053	0.055
intClust.3	203(26.5%)	87(7.2%)	**$2.1X10^{-32}$**	**$1.0X10^{-5}$**
intClust.4	191(25.0%)	152(12.5%)	**$1.2X10^{-12}$**	**$1.0X10^{-5}$**
intClust.5	21(2.7%)	168(13.9%)	**$2.6X10^{-16}$**	**$1.0X10^{-5}$**
intClust.6	27(3.5%)	59(4.9%)	0.155	4.03
intClust.7	92(12.0%)	97(8.0%)	**0.003**	**0.003**
intClust.8	156(20.4)	144(11.9%)	**$2.7X10^{-7}$**	**$1.0X10^{-5}$**
intClust.9	28(3.7%)	118(9.7%)	**$4.8X10^{-7}$**	**$1.0X10^{-5}$**
intClust.10	6(0.8%)	219(18.1%)	**$4.5X10^{-32}$**	**$1.0X10^{-5}$**

Bold = Statistically significant; HER2: human epidermal growth factor receptor 2; ER: oestrogen receptor; PgR: progesterone receptor;Triple negative: ER-/PgR-/HER2-. *Adjusted p values were calculated using Benjamini-Hochberg method to adjust for multiple testing.

Table 2: Multivariate analysis in the METABRIC cohort confirms that BLM mRNA over expression is a powerful independent prognostic factor.

	P-Value	**HR**	**95% CI for HR**	
			Lower	**Upper**
Breast Cancer Specific Survival				
BLM mRNA expression	**$2.0x10^{-6}$**	1.523	1.278	1.815
Size	**$1.0x10^{-6}$**	1.112	1.068	1.158
Grade				
G1		1.0		
G2	0.121	1.782	1.094	2.903
G3	**0.0044**	2.03	1.241	3.321
LN Status				
LN (1-3)	0.21	1.697	1.367	2.108
LN(>3)	**$1.0x10^{-6}$**	3.646	2.890	4.601

Bold: Statistically significant; HR: Hazard Ratio; CI: Confidence interval; LN: Lymph node

Table 3. BLM (nuclear and cytoplasmic protein co-expression) in breast cancer

VARIABLE	BLM Protein Expression				P- value	*P -Value (Adjusted)
	Nuc-/Cyto- (n= 332) N (%)	Nuc+/Cyto- (n=360) N (%)	Nuc-/Cyto+ (n=353) N (%)	Nuc+/Cyto+ (n=333) N (%)		
Tumour Grade						
G1	53 (16.0)	52 (21.8)	45 (12.9)	59 (17.7)	**$3.0X10^{-6}$**	**$1.0X10^{-5}$**
G2	87 (26.2)	208 (42.0)	102 (29.1)	108 (32.4)		
G3	192 (57.8)	86 (36.1)	203 (58.0)	166 (49.8)		
Mitotic Index						
M1 (low; mitoses < 10)	93 (28.4)	117 (49.4)	91 (26.1)	129 (38.9)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
M2 (medium; mitoses 10-18)	65 (19.8)	39 (16.5)	64 (18.3)	55 (16.6)		
M3 (high; mitosis >18)	170 (51.8)	81 (34.2)	194 (55.6)	148 (44.6)		
Pleomorphism						
1 (small-regular uniform)	12 (3.7)	6 (2.5)	2 (0.6)	8 (2.4)	**$1.2X10^{-5}$**	**$1.0X10^{-5}$**
2 (Moderate variation)	112 (34.1)	122 (51.5)	119 (34.2)	114 (34.4)		
3 (Marked variation)	204 (62.2)	109 (46.0)	227 (65.2)	209 (63.1)		
Tumour Type						
IDC-NST	170 (59.2)	105 (53.3)	204 (65.2)	170 (58.2)	**$6.6X10^{-5}$**	**$1.0X10^{-4}$**
Tubular Carcinoma	55 (19.2)	39 (19.8)	59 (18.8)	66 (22.6)		
Medullary Carcinoma	12 (4.2)	0 (0.0)	12 (3.8)	3 (1.0)		
ILC	28 (9.8)	30 (15.2)	17 (5.4)	18 (6.2)		
Others	22 (7.7)	23 (11.7)	21 (6.7)	35 (12.0)		
Triple Negative Phenotype						
No	244 (74.8)	210 (89.4)	248 (73.2)	285 (88.5)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
Yes	82 (25.2)	25 (10.6)	91 (26.8)	37 (11.5)		
ER						
Negative	110 (33.5)	40 (16.9)	112 (32.7)	68 (20.6)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
Positive	218 (66.5)	197 (83.1)	231 (67.3)	262 (79.4)		
BRCA1						
Absent	59 (24.6)	20 (13.2)	52 (20.3)	41 (17.3)	**0.036**	**0.047**
Normal	181 (75.4)	131 (86.8)	204 (79.7)	196 (82.7)		
XRCC1						
Low	61 (25.6)	23 (12.8)	27 (11.6)	153 (16.7)	**$1.7X10^{-4}$**	**$3.0X10^{-4}$**
High	177 (74.4)	156 (87.2)	205 (88.4)	761 (83.3)		
FEN1						
Low	192 (83.8) 37	117 (69.6)	169 (74.1)	152 (65.8)	**$1.0X10^{-4}$**	**$2.0X10^{-4}$**
High	(16.2)	51 (30.4)	59 (25.9)	79 (34.2)		
SMUG1						
Low	104 (47.1)	51 (33.3)	73 (34.4)	77 (35.5)	**0.013**	**0.018**
High	117 (52.9)	102 (66.7)	139 (65.6)	140 (64.5)		
APE1						
Low	185 (66.8)	93 (44.7)	99 (35.0)	532 (49.7)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	92 (33.2)	115 (55.3)	184 (65.0)	538 (50.3)		
Polβ						
Low	147 (50.9)	56 (25.9)	130 (42.1)	91 (30.6)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	142 (49.1)	160 (74.1)	179 (57.9)	206 (69.4)		
ATR						
Low	236 (75.9)	146 (69.5)	221 (67.4)	175 (55.6)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	75 (24.1)	64 (30.5)	107 (32.6)	140 (44.4)		
DNA-PKcs						
Low	126 (45.8)	58 (29.4)	124 (41.5)	68 (23.3)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	149 (54.2)	139 (70.6)	175 (58.5)	224 (76.7)		
MIB1						
Low	121 (44.5)	117 (57.6)	106 (37.7)	127 (44.9)	**$4.2X10^{-5}$**	**$1.0X10^{-4}$**
High	151 (55.5)	86 (42.4)	175 (62.3)	156 (55.1)		
P53						
Low expression	214 (78.1)	156 (85.2)	206 (72.0)	225 (80.9)	**0.005**	**0.008**
High expression	60 (21.9)	27 (14.8)	80 (28.0)	53 (19.1)		
Bcl-2						
Negative	119 (40.3)	56 (27.5)	127 (27.5)	99 (32.8)	**0.006**	**0.009**
Positive	176 (59.7)	148 (72.5)	148 (72.5)	203 (67.2)		
TOP2A						
Low	129 (56.6)	64 (39.8)	110 (43.1)	98 (41.4)	**0.001**	**0.002**
Overexpression	99 (43.4)	97 (60.2)	145 (56.9)	139 (58.6)		

Bold = statistically significant; BRCA1: Breast cancer 1, early onset; HER2: human epidermal growth factor receptor 2; ER: oestrogen receptor; PgR: progesterone receptor; CK: cytokeratin; Basal-like: ER-, HER2 and positive expression of either CK5/6, CK14 or EGFR; Triple negative: ER-/PgR-/HER2- . Adjusted p values were calculated using Benjamini-Hochberg false discovery rate method to adjust for multple testing. *Fischer test was used to obtain p values where one or more of cells has an expected frequency of five or less. For full data please also see supplementary Table S

FIGURE LEGENDS

Figure 1: Kaplan Meier curves showing BCSS (Breast cancer specific survival) based on BLM mRNA expression in **A.** whole cohort; **B.** ER+ cohort; **C.** ER- cohort; **D.** ER+ patients with NPI >3.4, who received endocrine therapy and **E.** ER- patients with NPI >3.4, who received chemotherapy.

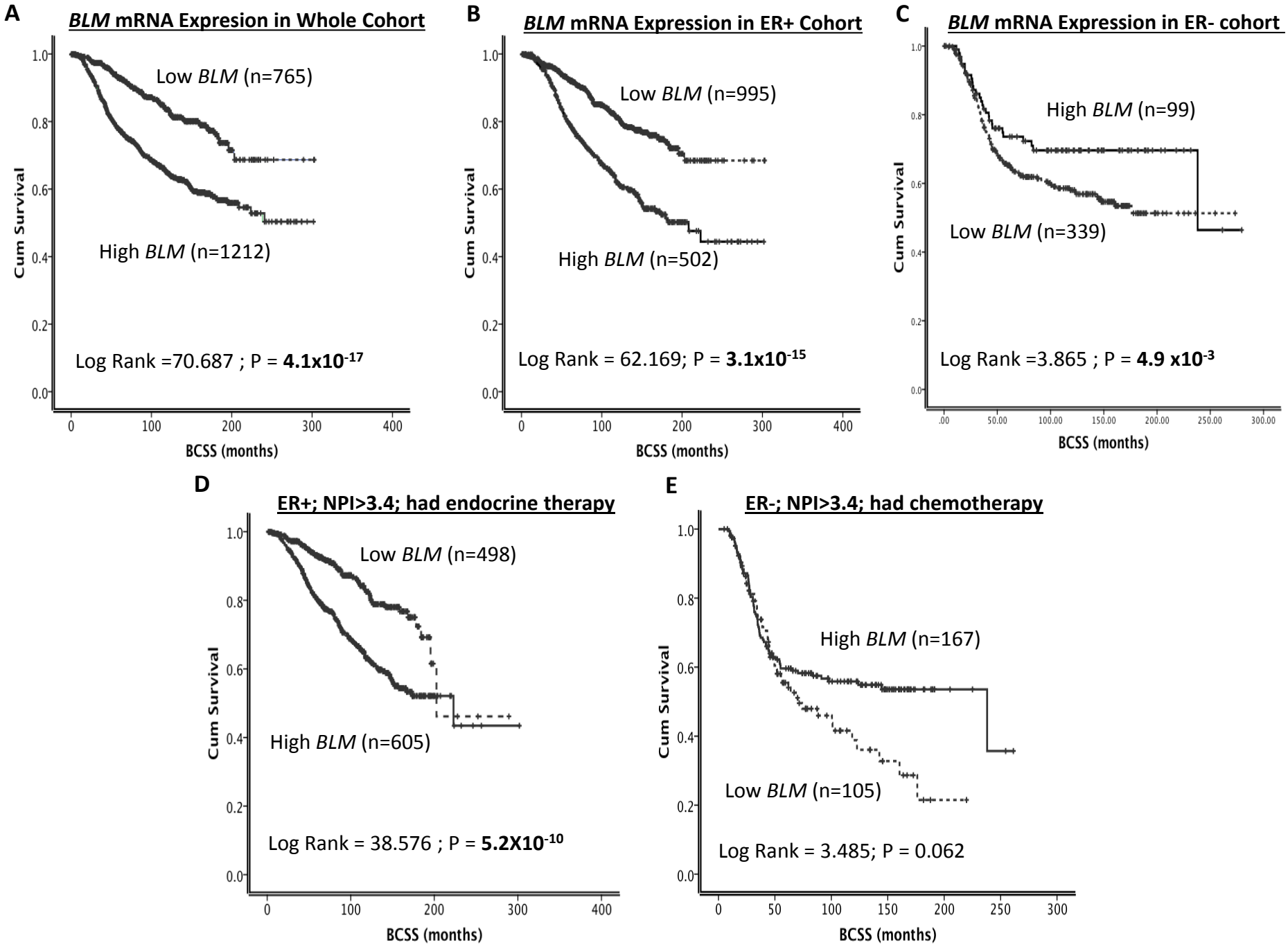

A
BLM mRNA Expresion in Whole Cohort
Low BLM (n=765)
High BLM (n=1212)
Log Rank =70.687 ; P = 4.1x10-17
Cum Survival
BCSS (months)
B
BLM mRNA Expression in ER+ Cohort
Low BLM (n=995)
High BLM (n=502)
Log Rank = 62.169; P = 3.1x10-15
Cum Survival
BCSS (months)
C
BLM mRNA Expression in ER- cohort
High BLM (n=99)
Low BLM (n=339)
Log Rank =3.865 ; P = 4.9 x10-3
Cum Survival
BCSS (months)
D
ER+; NPI>3.4; had endocrine therapy
Low BLM (n=498)
High BLM (n=605)
Log Rank = 38.576 ; P = 5.2X10-10
Cum Survival
BCSS (months)
E
ER-; NPI>3.4; had chemotherapy
High BLM (n=167)
Low BLM (n=105)
Log Rank = 3.485; P = 0.062
Cum Survival
BCSS (months)

Figure 1

Supplementary Table S1: Clinicopathological characteristics in the METABRIC cohort

Variables	N (%)
Age at diagnosis [Median (range)]	61.8 (21.93-96.29)
Tumour size [Median (range)]	23 (1, 182)
NPI [Median (95% CI)]	4.04 (3.99-4.09)
Survival [Median (Months, 95% Cl)]	149 (141-159)
Lymph nodes status	
0	1012
1	336
2	170
3	112
>3	316
ER status	
Positive	1485
Negative	437
PAM50 subtype	
Basal	322
HER2	238
Luminal A	714
Luminal B	484

Normal Not classified	188 6
<u>Adjuvant systemic therapy (AT</u>)	
No AT	290
Hormone therapy (HT)	1014
Chemotherapy	226
Hormone + chemotherapy	192

Supplementary Table S2: External validation cohorts (pooled n = 2413).

Study Code	Reference*	Number of patients	Number with Metastatic Relapse
Rosetta2002	Van de Vijver et al.,2002 [1]	295	101
PNAS1732912100	Sotiriou et al., 2003 [2]	99	30
GSE2603	Minn et al., 2005 [3]	82	27
GSE1456	Pawitan et al., 2005 [4]	159	40
GSE2034	Wang et al., 2005 [5]	286	107
GSE2741	Weigelt et al., 2005 [6]	88	20
E_TABM_	Chin et al., 2006 [7]	112	21
GSE7390	Desmedt et al., 2007 [8]	198	62
GSE6532	Loi et al., 2007 [9]	393	101
GSE5327	Minn et al., 2007 [10]	58	11
GSE7849	Anders et al., 2008 [11]	75	14
GSE9893	Chanrion et al., 2008 [12]	155	48
GSE9195	Loi et al., 2008 [13]	77	10
GSE11121	Schmidt et al., 2008 [14]	200	46
GSE12093	Zhang et al., 2009 [15]	136	20
Total:		2413	658

*** References**

1. Van De Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DWet al. A gene-expression signature as a predictor of survival in breast cancer. New England Journal of Medicine. 2002;347(25):1999-2009.

2. Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri Aet al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proceedings of the National Academy of Sciences. 2003;100(18):10393-8.

3. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DDet al. Genes that mediate breast cancer metastasis to lung. Nature. 2005;436(7050):518-24.

4. Pawitan Y, Bjöhle J, Amler L, Borg A-L, Egyhazi S, Hall Pet al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Research. 2005;7(6):R953.

5. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang Fet al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. The Lancet. 2005;365(9460):671-9.

6. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MGet al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer research. 2005;65(20):9155-8.

7. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo W-Let al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer cell. 2006;10(6):529-41.

8. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains Bet al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clinical cancer research. 2007;13(11):3207-14.

9. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet Cet al. Definition of clinically distinct molecular subtypes in estrogen receptor–positive breast carcinomas through genomic grade. Journal of clinical oncology. 2007;25(10):1239-46.

10. Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten Det al. Lung metastasis genes couple breast tumor size and metastatic spread. Proceedings of the National Academy of Sciences. 2007;104(16):6740-5.

11. Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, Foekens JAet al. Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. PloS one. 2008;3(1):e1373.

12. Chanrion M, Negre V, Fontaine H, Salvetat N, Bibeau F, Mac Grogan Get al. A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. Clinical Cancer Research. 2008;14(6):1744-52.

13. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AMet al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC genomics. 2008;9(1):239.

14. Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch Het al. The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer research. 2008;68(13):5405-13.

15. Zhang Y, Sieuwerts AM, McGreevy M, Casey G, Cufer T, Paradiso Aet al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. Breast cancer research and treatment. 2009;116(2):303-9.

Supplementary Table S3: Clinicopathological characteristics of Nottingham cohort

Variable	n*	Cases	(%)
Menopausal status	1650		
Pre-menopausal		612	(37.0)
postmenopausal		1038	(63.0)
Tumour Grade (NGS)	1650		
G1		306	(18.5)
G2		531	(32.2)
G3		813	(49.3)
Lymph node stage	1650		
Negative		1056	(64.0)
Positive (1-3 nodes)		486	(29.5)
Positive (>3 nodes)		108	(6.5)
Tumour size (cm)	1650		
T1 a + b (≤1.0)		187	(11.0)
T1 c (>1.0 -2.0)		868	(53.0)
T2 (>2.0-5)		579	(35.0)
T3 (>5)		16	(1.0)
Tumour type	1650		
IDC-NST		941	(57)
Tubular		349	(21)
ILC		160	(10)
Medullary (typical/atypical)		41	(2.5)
Others		159	(9.5)
NPI subgroups	1650		

Excellent PG(2.08-2.40)	Low risk	207	(12.5)
Good PG(2.42-3.40)		331	(20.1)
Moderate I PG(3.42 to 4.4)	High risk	488	(29.6)
Moderate II PG(4.42 to 5.4)		395	(23.9)
Poor PG(5.42 to 6.4)		170	(10.3)
Very poor PG(6.5–6.8)		59	(3.6)
Survival at 20 years	1650		
Alive and well		1055	(64.0)
Dead from disease		468	(28.4)
Dead from other causes		127	(7.6)
Adjuvant systemic therapy (AT)			
No AT		665	(42.0)
Hormone therapy (HT)		642	(41.0)
Chemotherapy		307	(20.0)
Hormone + chemotherapy		46	(3.0)

* Number of cases for which data were available.

NPI; Nottingham prognostic index, PG; prognostic group

Supplementary Table S4: Antigens, primary antibodies, clone, source, optimal dilution and scoring system used for each immunohistochemical marker

Antigen	Antibody	Clone	Source	Antigen Retrieval	Dilution / Incubation Time	Distribution	Scoring system	Cut-offs
BRCA1	BRCA1	MS110	Calbiochem	Citrate pH6	1:100 60 min	Nuclear	% of positive cells	<25% (negative)
ATR	Mouse MAb Anti-ATR	1E9	Novus Biologicals	Citrate pH6	1:20 18 hours	Nuclear	H-score	≥60 (High)
pChk1	Rabbit anti-pChk1	Ab58567	Abcam	Citrate pH6	1:140 60 min	Nuclear	H-score	≥50 (High)
DNA-PKcs	Mouse MAb Anti-	3H6	Abcam	Citrate pH6	1:1000 20 min	Nuclear	H-score	>260 (high)
XRCC1	Mouse MAb Anti-XRCC1	33-2-5	Thermo-scientific	Citrate pH6	1:200 20 min	Nuclear	% of positive cells	≥10% (positive)
APE1	Rabbit anti-APE-1	polyclonal	Novus Biologicals	Citrate pH6	1:500 60 min	Nuclear	H-score	>100 (positive)
SMUG1	Goat anti-SMUG1	polyclonal	Acris Antibodies	Citrate pH6	1:200 15 min	Nuclear	H-score	≤35 (negative)

FEN1	Rabbit anti-FEN1	polyclonal	Novus Biologicals	Citrate pH6	1:200 60 min	Nuclear and Cytoplasm	H-score	≤100 (negative)
Rad51	Mouse anti-Rad51	polyclona	Abcam	Citrate pH6	1:70 60 min	Nuclear	H-score	≥10 (positive)
P21	Mouse MAb anti-p21	SW118	Dako-Cytomation	Citrate pH6	1:50 60 min	Nuclear	% of positive cells	≥10% (positive)
Ki67	Mouse MAb anti-Ki-67	MIB1	Dako-Cytomation	Citrate pH6	1:300 60 min	Nuclear	% of positive cells	< 10% (low) 10-30% (moderate) >30% (high)
P53	Mouse MAb anti p53	DO7	Novocastra	Citrate pH6	1: 50 60 min	Nuclear	% of positive cells	≤20% (negative) >20% (High)
Bcl-2	Mouse MAb anti-Bcl2	124	Dako-Cytomation	Citrate pH6	1:100 60 min	Cytoplasm	% of positive cells	>10% (positive
ER	Mouse MAb anti-ER-α	SP1	Dako-Cytomation	Citrate pH6	1:150 30 min	Nuclear	Allred score	≥3 (positive)
ER	Mouse MAb anti-ER-α	EP1	Dako-Cytomation	Citrate pH6	1:80 30 min	Nuclear	% positive cells	≥1% positive
PR	Mouse MAb anti-PR	PgR636	Dako-Cytomation	Citrate pH6	1:125 30 min	Nuclear	% positive cells	≥1% positive

CK14	Mouse MAb anti-Ck14	LL002	Novocastra	Citrate pH6	1:40 60 min	Cytoplasm	% of positive cells	≥10% (positive)
Ck5/6	Mouse MAb anti-Ck5/6	D5/161B4	Dako-Cytomation	EDTA pH8	1:100 60 min	Cytoplasm	% of positive cells	≥10% (positive)
Ck17	Mouse MAb anti-Ck17	E3	Dako-Cytomation	Citrate pH6	1:100 60 min	Cytoplasm	% of positive cells	≥10% (positive)
Ck18	Mouse MAb anti-Ck18	DC10	Dako-Cytomation	Citrate pH6	1:100 60 min	Cytoplasm	% of positive cells	≥10% (positive)
HER2	Rabbit antihuman c-erbB2	polyclonal	Dako-Cytomation	None	1:400 60 min	Membrane	See text	See text
TOP2A	Mouse MAb	KiS1	Dako-Cytomation	Citrate pH6	1:150 60 min	Nuclear/ cytoplasm	% of positive cells	>25% (positive)

All sections were pre-treated with microwave antigen retrieval using 0.1% citrate buffer (pH 6) except for HER2 (no pre-treatment) and EGFR (pre-treated with protease for 10 minutes).

Supplementary Table S5: BLM (nuclear protein expression) in breast cancer

VARIABLE	BLM (Nuclear) Protein Expression		Unadjusted P- values	*Adjusted P -Values
	Low (n=683) N (%)	High (n=574) N (%)		
A) Pathological Parameters				
Tumour Size				
<1cm	50 (7.3)	67 (11.7)	**0.012**	**0.018**
>1-2cm	345 (50.6)	283 (49.6)		
>2-5cm	265 (38.9)	213 (37.3)		
>5cm	22 (3.2)	8 (1.4)		
Tumour Stage				
1	422 (61.8)	356 (62.2)	0.874	36.70
2	198 (29.0)	168 (29.4)		
3	63 (9.2)	48 (8.4)		
Tumour Grade				
G1	98 (14.4)	111 (19.4)	**$7.0X10^{-6}$**	**$1.0X10^{-5}$**
G2	189 (27.7)	208 (36.4)		
G3	395 (57.9)	252 (44.1)		
Mitotic Index				
M1 (low; mitoses < 10)	184 (27.2)	246 (43.2)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
M2 (medium; mitoses 10-18)	129 (19.1)	94 (16.5)		
M3 (high; mitosis >18)	364 (53.8)	229 (40.2)		
Tubule Formation				
1 (>75% of definite tubule)	34 (5.0)	34 (6.0)	0.348	0.365
2 (10%-75% definite tubule)	219 (32.3)	201 (35.3)		
3 (<10% definite tubule)	424 (62.6)	334 (58.7)		
Pleomorphism				
1 (small-regular uniform)	14 (2.1)	12 (1.5)	**0.020**	**0.029**
2 (Moderate variation)	231 (34.2)	236 (41.5)		
3 (Marked variation)	431 (63.8)	318 (56.0)		
Tumour Type				
IDC-NST	374 (62.3)	275 (56.2)	**$1.2X10^{-4}$**	**$4.0X10^{-4}$**
Tubular Carcinoma	114 (19.0)	105 (21.5)		
Medullary Carcinoma	24 (4.0)	3 (0.6)		
ILC	45 (7.5)	48 (9.8)		
Others	43 (7.2)	58 (11.9)		
Lymphovascular Invasion				
No	454 (67.6)	362 (64.0)	0.183	0.207
Yes	218 (32.4)	204 (36.0)		
B) Aggressive phenotype				

Her2 overexpression No Yes	 603 (89.9) 68 (10.1)	 487 (87.7) 68 (12.3)	 0.240	 0.265
Triple Negative Phenotype No Yes	 492 (74.0) 173 (26.0)	 495 (88.9) 62 (11.1)	 **$1.0X10^{-6}$**	 **$1.0X10^{-5}$**
Basal Like Phenotype No Yes	 527 (82.7) 110 (17.3)	 503 (92.0) 44 (8.0)	 **$3.0X10^{-6}$**	 **$1.0X10^{-5}$**
Cytokeratin 6 (CK6) Negative Positive	 475 (80.4) 116 (19.6)	 416 (88.1) 56 (11.9)	 **0.001**	 **0.002**
Cytokeratin 14 (CK14) Negative Positive	 499 (85.3) 86 (14.7)	 419 (89.7) 48 (10.3)	 **0.033**	 **0.042**
Cytokeratin 18 (CK18) Negative Positive	 84 (15.4) 460 (84.6)	 19 (4.4) 411 (95.6)	 **$1.0X10^{-6}$**	 **$1.0X10^{-5}$**
Cytokeratin 19 (CK19) Negative Positive	 46 (7.8) 545 (92.2)	 20 (4.3) 446 (95.7)	 **0.020**	 **0.028**
C) Hormone receptors				
ER Negative Positive	 222 (33.1) 449 (66.9)	 108 (19.0) 459 (81.0)	 **$1.0X10^{-6}$**	 **$1.0X10^{-5}$**
PgR Negative Positive	 307 (47.8) 335 (52.2)	 202 (38.5) 323 (61.5)	 **0.001**	 **0.002**
AR Negative Positive	 248 (44.5) 309 (55.5)	 123 (28.0) 316 (72.0)	 **$1.0X10^{-6}$**	 **$1.0X10^{-5}$**
D) DNA Repair				
BRCA1 Absent Normal	 111 (22.4) 385 (77.6)	 61 (15.7) 327 (84.3)	 **0.013**	 **0.019**

XRCC1 Low High	103 (20.5) 400 (79.5)	50 (12.2) 361 (87.8)	**0.001**	**0.002**
FEN1 Low High	361 (79.0) 96 (21.0)	269 (67.4) 130 (32.6)	**$1.0X10^{-5}$**	**$3.0X10^{-4}$**
SMUG1 Low High	177 (40.9) 256 (59.1)	128 (34.6) 242 (65.4)	0.067	0.082
APE1 Low High	340 (58.7) 239 (41.3)	192 (39.1) 299 (60.9)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
PolB Low High	277 (46.3) 321 (53.7)	147 (28.7) 366 (71.3)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
ATR Low High	35 (6.1) 538 (93.9)	18 (3.9) 449 (96.1)	**$1.8X10^{-4}$**	**$5.0X10^{-4}$**
ATM Low High	223 (52.0) 206 (48.0)	179 (53.8) 154 (46.2)	0.627	0.642
DNA-PK Low High	250 (43.6) 324 (56.4)	126 (25.8) 363 (74.2)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
<u>E) Cell cycle/apoptosis regulators</u>				
P16 Low High	396 (80.8) 94 (19.2)	347 (93.8) 23 (6.2)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
P21 Low High	316 (60.5) 206 (39.5)	202 (53.2) 178 (46.8)	**0.027**	**0.036**
MIB1 Low High	227 (41.0) 326 (59.0)	244 (50.2) 242 (49.8)	**0.003**	**0.006**

P53 Low expression High expression	420 (75.0) 140 (25.0)	381 (82.6) 80 (17.4)	**0.003**	**0.005**
Bcl-2 Negative Positive	246 (40.2) 366 (59.8)	155 (30.6) 351 (69.4)	**0.001**	**0.002**
TOP2A Low Overexpression	239 (49.5) 244 (50.5)	162 (40.7) 236 (59.3)	**0.009**	**0.015**
pCHK1 (Nuclear) Low High	616 (90.2) 67 (9.8)	415 (72.3) 159 (27.7)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
pCHK1 (Cytoplasmic) Low High	191 (28.0) 492 (72.0)	124 (21.6) 450 (78.4)	**0.010**	**0.016**
Non-phospho CHK1 (Cyto.) Low High	284 (52.0) 262 (48.0)	205 (45.2) 249 (54.8)	**0.031**	**0.041**
CHK2 Low High	258 (50.5) 253 (49.5)	164 (41.0) 236 (59.0)	**0.004**	**0.007**
Bax Low High	272 (68.9) 123 (31.1)	235 (72.1) 91 (27.9)	0.345	0.371
CDK1 Low High	303 (67.2) 148 (32.8)	247 (72.0) 96 (28.0)	0.144	0.172
CDK18 (Cytoplasmic) Low High	426 (78.5) 117 (21.5)	318 (70.7) 132 (29.3)	**0.005**	**0.008**
RECQL5 Low High	295 (56.0) 232 (44.0)	167 (36.9) 285 (63.1)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
MDM2 Low Overexpression	386 (77.4) 113 (22.6)	272 (73.3) 99 (26.7)	0.170	0.198

Bold = Statistically significant; BRCA1: Breast cancer 1, early onset; HER2: human epidermal growth factor receptor 2; ER: oestrogen receptor; PgR: progesterone receptor; CK: cytokeratin;

Basal-like: ER-, HER2 and positive expression of either CK5/6, CK14 or EGFR; Triple negative: ER-/PgR-/HER2- . *Adjusted p values were calculated using Benjamini-Hochberg method to adjust for multiple testing.

Supplementary Table S6: BLM (cytoplasmic protein expression) in breast cancer

VARIABLE	BLM (Cytoplasmic) Protein Expression		Unadjusted P-Values	*Adjusted P-Values
	Low (n=571) N (%)	High (n=686) N (%)		
A) Pathological Parameters				
Tumour Size				
<1cm	59 (6.9)	69 (8.5)	0.204	0.343
>1-2cm	274 (48.1)	325 (51.8)		
>2-5cm	219 (38.4)	235 (37.9)		
>5cm	18 (3.2)	13 (1.8)		
Tumour Stage				
1	360 (63.0)	418 (61.1)	0.545	0.789
2	158 (27.7)	181 (30.4)		
3	53 (9.3)	62 (8.5)		
Tumour Grade				
G1	105 (18.4)	104 (15.2)	0.137	0.274
G2	187 (32.8)	210 (30.7)		
G3	278 (48.8)	369 (54.0)		
Mitotic Index				
M1 (low; mitoses < 10)	210 (31.3)	220 (36.5)	0.108	0.238
M2 (medium; mitoses 10-18)	104 (17.6)	119 (19.4)		
M3 (high; mitosis >18)	251 (51.5)	342 (44.0)		
Tubule Formation				
1 (>75% of definite tubule)	31 (5.5)	37 (5.4)	0.998	41.91
2 (10%-75% definite tubule)	190 (33.6)	230 (33.8)		
3 (<10% definite tubule)	344 (60.9)	414 (60.8)		
Pleomorphism				
1 (small-regular uniform)	18 (3.2)	10 (1.5)	**0.002**	**0.012**
2 (Moderate variation)	234 (41.4)	233 (34.3)		
3 (Marked variation)	313 (55.4)	436 (64.2)		
Tumour Type				
IDC-NST	275 (56.8)	374 (61.8)	**0.009**	**0.034**
Tubular Carcinoma	94 (19.4)	125 (20.7)		
Medullary Carcinoma	12 (2.5)	15 (2.5)		
ILC	58 (12.0)	35 (5.8)		
Others	45 (9.3)	56 (9.3)		
Lymphovascular Invasion				
No	363 (65.1)	453 (66.6)	0.564	0.789
Yes	195 (34.9)	227 (33.4)		
B) Aggressive phenotype				

Her2 overexpression No Yes	 495 (88.9) 62 (11.1)	 595 (88.9) 74 (11.1)	0.969	1.02
Triple Negative Phenotype No Yes	 454 (80.9) 107 (19.1)	 533 (80.6) 128 (19.4)	0.897	0.99
Basal Like Phenotype No Yes	 476 (87.8) 66 (12.2)	 554 (86.3) 88 (13.7)	0.436	0.704
Cytokeratin 6 (CK6) Negative Positive	 391 (83.4) 78 (16.6)	 500 (84.2) 94 (15.8)	0.723	0.893
Cytokeratin 14 (CK14) Negative Positive	 396 (85.7) 66 (14.3)	 522 (88.5) 68 (11.5)	0.183	0.334
Cytokeratin 18 (CK18) Negative Positive	 55 (12.5) 385 (87.5)	 48 (9.0) 486 (91.0)	0.076	0.187
Cytokeratin 19 (CK19) Negative Positive	 35 (7.4) 436 (92.6)	 31 (5.3) 555 (94.7)	0.153	0.292
C) Hormone receptors				
ER Negative Positive	 150 (26.5) 415 (73.5)	 180 (26.7) 493 (73.3)	0.938	1.01
PgR Negative Positive	 238 (44.2) 300 (55.8)	 271 (43.1) 358 (56.9)	0.692	0.880
AR Negative Positive	 170 (38.3) 274 (61.7)	 201 (36.4) 351 (63.6)	0.543	0.884
D) DNA Repair				
BRCA1 Absent Normal	 79 (20.2) 312 (79.8)	 93 (18.9) 400 (81.1)	0.617	0.835

XRCC1 Low High	84 (20.1) 333 (79.9)	69 (13.9) 428 (86.1)	**0.012**	**0.039**
FEN1 Low High	309 (77.8) 88 (22.2)	321 (69.9) 138 (30.1)	**0.009**	**0.031**
SMUG1 Low High	155 (41.4) 219 (58.6)	150 (35.0) 279 (65.0)	0.059	0.154
APE1 Low High	278 (57.3) 207 (42.7)	254 (43.4) 331 (56.6)	**$6.0X10^{-6}$**	**$1.0X10^{-5}$**
PolB Low High	203 (40.2) 302 (59.8)	221 (36.5) 385 (63.5)	0.203	0.355
ATR Low High	382 (73.3) 139 (26.7)	315 (61.6) 204 (38.4)	**$2.4X10^{-5}$**	**0.0003**
ATM Low High	182 (53.4) 159 (46.6)	220 (52.3) 201 (47.7)	0.759	0.884
DNA-PK Low High	184 (39.0) 288 (61.0)	192 (32.5) 399 (67.5)	**0.028**	0.084
E) Cell cycle/apoptosis regulators				
P16 Low High	338 (86.4) 53 (13.6)	405 (86.4) 64 (13.6)	0.969	0.992
P21 Low High	235 (57.5) 174 (42.5)	283 (57.4) 210 (42.6)	0.76	0.86
MIB1 Low High	238 (50.1) 237 (49.9)	233 (41.3) 331 (58.7)	**0.005**	**0.021**

P53 Low expression High expression	370 (81.0) 87 (19.0)	431 (76.4) 133 (23.6)	0.079	0.184
Bcl-2 Negative Positive	175 (35.1) 324 (64.9)	226 (36.5) 393 (63.5)	0.618	0.811
TOP2A Low Overexpression	193 (49.6) 196 (50.4)	208 (42.3) 284 (57.7)	**0.030**	0.084
pCHK1 (Nuclear) Low High	458 (80.2) 113 (19.8)	573 (83.5) 113 (16.5)	0.127	0.266
pCHK1 (Cytoplasmic) Low High	187 (32.7) 384 (67.3)	128 (18.7) 558 (81.3)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
Non-phospho CHK1 (Cyto.) Low High	243 (54.1) 206 (45.9)	246 (44.6) 305 (55.4)	**0.003**	**0.015**
CHK2 Low High	217 (54.1) 184 (45.9)	205 (40.2) 305 (59.8)	**$2.9X10^{-5}$**	**$2.0X10^{-4}$**
Bax Low High	237 (76.0) 75 (24.0)	270 (66.0) 139 (34.0)	**0.004**	**0.018**
CDK1 Low High	240 (70.6) 100 (29.4)	310 (68.3) 144 (31.7)	0.486	0.756
CDK18 (Cytoplasmic) Low High	367 (81.7) 82 (18.3)	377 (69.3) 167 (30.7)	**$7.0X10^{-6}$**	**$1.0X10^{-4}$**
RECQL5 Low High	235 (53.5) 204 (46.5)	227 (42.0) 313 (58.0)	**$3.4X10^{-4}$**	**$2.4X10^{-3}$**
MDM2 Low Overexpression	296 (75.1) 98 (24.9)	362 (76.1) 114 (23.9)	0.752	0.902

Bold = Statistically significant; BRCA1: Breast cancer 1, early onset; HER2: human epidermal growth factor receptor 2; ER: oestrogen receptor; PgR: progesterone receptor; CK: cytokeratin;

Basal-like: ER-, HER2 and positive expression of either CK5/6, CK14 or EGFR; Triple negative: ER-/PgR-/HER2- . *Adjusted p values were calculated using Benjamini-Hochberg method to adjust for multiple testing.

Supplementarty Table S7. BLM (nuclear and cytoplasmic protein co-expression) in breast cancer

VARIABLE	BLM Protein Expression				P- value	*P -Value (Adjusted)
	Nuc-/Cyto- (n= 332) N (%)	Nuc+/Cyto- (n=360) N (%)	Nuc-/Cyto+ (n=353) N (%)	Nuc+/Cyto+ (n=333) N (%)		
A) Pathological Parametersic						
Tumour Size						
<1cm	29 (8.7)	30 (12.6)	21 (6.0)	37 (11.1)	0.065	0.083
>1-2cm	163 (49.1)	111 (46.6)	182 (52.0)	172 (51.7)		
>2-5cm	127 (38.3)	92 (38.7)	138 (39.4)	121 (36.3)		
>5cm	13 (3.9)	5 (2.1)	9 (2.6)	3 (0.9)		
Tumour Stage						
1	207 (62.3)	153 (64.0)	215 (61.3)	203 (61.0)	0.946	39.73
2	92 (27.7)	66 (27.6)	106 (30.2)	102 (30.6)		
3	33 (9.9)	20 (8.4)	30 (8.5)	28 (8.4)		
Tubule Formation						
1 (>75% definite tubule)	17 (5.2)	14 (5.9)	17 (4.9)	20 (6.0)	0.90	0.92
2 (10%-75% definite tubule)	107 (32.6)	83 (35.0)	112 (32.1)	118 (35.5)		
3 (<10% definite tubule)	204 (62.2)	140 (59.1)	220 (63.0)	194 (58.4)		
Pleomorphism						
1 (small-regular uniform)	12 (3.7)	6 (2.5)	2 (0.6)	8 (2.4)	**$1.2X10^{-5}$**	**$1.0X10^{-5}$**
2 (Moderate variation)	112 (34.1)	122 (51.5)	119 (34.2)	114 (34.4)		
3 (Marked variation)	204 (62.2)	109 (46.0)	227 (65.2)	209 (63.1)		

Lymphovascular Invasion						
No	219 (67.2)	144 (62.1)	235 (67.9)	218 (65.3)	0.486	0.551
Yes	107 (32.8)	88 (37.9)	111 (32.1)	116 (34.7)		
B) Aggressive phenotype						
Her2 overexpression						
No	290 (89.2)	205 (88.4)	313 (90.5)	282 (87.3)	0.617	0.664
Yes	35 (10.8)	27 (11.6)	33 (9.5)	41 (12.7)		
Basal Like Phenotype						
No	260 (83.3)	216 (93.9)	267 (82.2)	287 (90.5)	**$2.9X10^{-5}$**	**$1.0X10^{-4}$**
Yes	52 (16.7)	14 (6.1)	58 (17.8)	30 (9.5)		
Cytokeratin 6 (CK6)						
Negative	223 (79.6)	168 (88.9)	252 (81.0)	248 (87.6)	**0.007**	**0.011**
Positive	57 (20.4)	21 (11.1)	59 (19.0)	35 (12.4)		
Cytokeratin 18 (CK18)						
Negative	49 (18.6)	6 (3.4)	35 (12.5)	13 (5.1)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
Positive	215 (81.4)	170 (96.6)	245 (87.5)	241 (94.9)		
Cytokeratin 19 (CK19)						
Negative	29 (10.2)	6 (3.2)	17 (5.5)	14 (5.0)	**0.008**	**0.012**
Positive	254 (89.8)	182 (96.8)	291 (94.5)	264 (95.0)		
C) Hormone receptors						
PR						
Negative	151 (47.6)	87 (39.4)	156 (48.0)	115 (37.8)	**0.016**	**0.022**
Positive	166 (52.4)	134 (60.6)	169 (52.0)	189 (62.2)		
AR						
Negative	126 (47.0)	44 (25.0)	122 (42.2)	79 (30.0)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
Positive	142 (53.0)	316 (75.0)	167 (57.6)	184 (70.0)		

D) DNA Repair						
ATM						
Low	109 (54.0)	73 (52.5)	114 (50.2)	106 (54.6)	0.806	0.846
High	93 (46.0)	66 (47.5)	113 (49.8)	88 (45.4)		
E) Cell cycle/apoptosis regulators						
P16						
Low	199 (81.9)	139 (93.9)	197 (79.8)	208 (93.7)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	44 (18.1)	9 (6.1)	50 (20.2)	14 (6.3)		
pCHK1 (Nuclear)						
Low	298 (90.0)	160 (66.7)	318 (90.3)	255 (76.3)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	33 (10.0)	80 (33.3)	34 (9.7)	79 (23.7)		
pCHK1 (Cytoplasmic)						
Low	123 (37.2)	64 (26.7)	68 (19.3)	60 (18.0)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	208 (62.8)	176 (73.3)	284 (80.7)	274 (82.0)		
Non-phospho CHK1						
Low	151 (57.0)	92 (50.0)	133 (47.3)	113 (41.9)	**0.005**	**0.008**
High	114 (43.0)	92 (50.0)	148 (52.7)	157 (58.1)		
CHK2						
Low	145 (59.7)	72 (45.6)	113 (42.2)	92 (38.0)	**$9.0X10^{-6}$**	**$1.0X10^{-5}$**
High	98 (40.3)	86 (54.4)	155 (57.8)	150 (62.0)		
Bax						
Low	138 (75.4)	99 (76.7)	134 (63.2)	136 (69.0)	**0.018**	**0.024**
High	123 (24.6)	30 (23.3)	78 (36.8)	61 (31.0)		
CDK18 (Cytoplasmic)						
Low	223 (84.8)	144 (77.4)	203 (72.5)	174 (65.9)	**$7.0X10^{-6}$**	**$1.0X10^{-5}$**
High	40 (15.2)	42 (22.6)	77 (27.5)	90 (34.1)		
RECQL5						

Low	161 (63.6)	74 (39.8)	134 (48.9)	93 (35.0)	**$1.0X10^{-6}$**	**$1.0X10^{-5}$**
High	92 (36.4)	112 (60.2)	140 (51.1)	173 (65.0)		
MDM2						
Low	184 (76.3)	112 (73.2)	202 (78.3)	160 (73.4)	0.544	0.601
Overexpression	57 (23.7)	41 (26.8)	56 (21.7)	58 (26.6)		

Bold = statistically significant; BRCA1: Breast cancer 1, early onset; HER2: human epidermal growth factor receptor 2; ER: oestrogen receptor; PgR: progesterone receptor; CK: cytokeratin; Basal-like: ER-, HER2 and positive expression of either CK5/6, CK14 or EGFR; Triple negative: ER-/PgR-/HER2- . Adjusted p values were calculated using Benjamini-Hochberg false discovery rate method to adjust for multple testing. *Fischer test was used to obtain p values where one or more of cells has an expected frequency of five or less.

Supplementary Table S8. BLM - Rad51 nuclear co-expression and breast cancer

VARIABLE	BLM-Rad51(Nuclear) Protein Co-Expression				P- value	P -Value (Adjusted)
	BLM-/Rad51- (n=107) N (%)	BLM+ /Rad51- (n=295) N (%)	BLM- /Rad51+ (n=88) N (%)	BLM+ /Rad51+ (n=273) N (%)		
A) Pathological Parameters						
Tumour Size						
≤1cm	6 (5.6)	18 (8.0)	4 (5.6)	26 (9.8)	**0.004**	**0.008**
>1-2cm	49 (45.8)	97 (43.1)	49 (69.0)	138 (52.3)		
>2-5cm	49 (45.8)	106 (49.1)	16 (22.5)	99 (37.5)		
>5cm	3 (2.8)	4 (1.8)	2 (2.8)	1 (0.4)		
Tumour Stage						
1	65 (60.7)	116 (51.3)	46 (63.9)	153 (58.0)	0.251	0.326
2	30 (28.0)	88 (38.9)	23 (31.9)	85 (32.2)		
3	12 (11.2)	22 (9.7)	3 (4.2)	26 (9.8)		
Tumour Grade						
G1	9 (16.0)	27 (11.9)	9 (12.7)	48 (18.2)	**$7.0X10^{-5}$**	**$3.0X10^{-4}$**
G2	21 (26.2)	70 (31.0)	22 (31.0)	103 (39.0)		
G3	77 (57.8)	129 (57.1)	40 (56.3)	113 (42.8)		
Mitotic Index						
M1 (low; mitoses < 10)	16 (15.2)	65 (29.7)	19 (27.1)	103 (39.8)	**$1.0X10^{-4}$**	**$3.0X10^{-4}$**
M2 (medium; mitoses 10-18)	19 (18.1)	42 (19.2)	10 (14.3)	51 (19.7)		
M3 (high; mitosis >18)	70 (66.7)	112 (51.1)	41 (58.6)	105 (40.5)		
Tubule Formation						
1 (>75% definite tubule)	2 (1.9)	9 (4.1)	3 (4.3)	12 (4.6)	0.366	4.75
2 (10%-75% definite tubule)	30 (28.6)	65 (35.0)	24 (34.3)	97 (37.5)		
3 (<10% definite tubule)	73 (69.5)	145 (66.2)	43 (61.4)	150 (57.9)		
Pleomorphism						
1 (small-regular uniform)	0 (0.0)	0 (0.0)	1 (1.4)	4 (1.6)	**0.011**	**0.02**
2 (Moderate variation)	26 (25.0)	78 (35.6)	24 (34.3)	111 (43.0)		
3 (Marked variation)	78 (75.0)	141 (64.4)	45 (64.3)	143 (55.4)		
Tumour Type						
IDC-NST	81 (75.7)	142 (62.8)	47 (65.3)	139 (52.7)	**0.040**	**0.06**
Tubular Carcinoma	8 (7.5)	37 (16.4)	13 (18.1)	69 (26.2)		
Medullary Carcinoma	7 (6.5)	5 (2.2)	1 (1.4)	3 (1.2)		
ILC	4 (3.8)	23 (10.1)	6 (8.4)	27 (10.3)		
Others	7 (6.5)	23 (9.5)	21 (6.8)	35 (9.6)		
Lymph Node Status						
Negative	52 (61.9)	198 (50.3)	38 (61.3)	135 (56.3)	0.320	0.34
Positive (1-3)	26 (31.0)	80 (41.0)	23 (37.1)	88 (36.7)		
Positive (>3)	6 (7.1)	17 (8.7)	1 (1.6)	17 (7.1)		

B) Aggressive Phenotype						
Her2 overexpression No Yes	97 (90.7) 10 (9.3)	184 (82.9) 38 (17.1)	62 (87.3) 9 (12.7)	217 (84.4) 40 (15.6)	0.278	0.319
Triple Negative Phenotype No Yes	88 (82.2) 19 (17.8)	191 (84.5) 35 (15.5)	54 (75.0) 18 (25.0)	228 (86.4) 36 (13.6)	0.129	0.18
NPI ≤3.4 >3.4	14 (14.1) 85 (85.9)	49 (22.6) 168 (77.4)	19 (27.5) 50 (72.5)	86 (34.0) 167 (66.0)	**0.001**	**0.0026**
C) Hormone Receptors						
ER Negative Positive	49 (47.1) 55 (52.9)	67 (30.5) 153 (69.5)	21 (30.9) 47 (69.1)	46 (17.9) 211 (82.1)	**$3.8X10^{-7}$**	**$1.0X10^{-5}$**
PR Negative Positive	65 (64.4) 36 (35.6)	102 (47.9) 111 (52.1)	30 (42.9) 40 (57.1)	89 (35.3) 163 (64.7)	**$1.0X10^{-5}$**	**$1.0X10^{-4}$**

Bold = Statistically significant; HER2: human epidermal growth factor receptor 2; ER: oestrogen receptor; PgR: progesterone receptor; Triple negative: ER-/PgR-/HER2- . *Adjusted p values were calculated using Benjamini-Hochberg false discovery rate method to adjust for multple testing.

Supplementary Table S9: Multivariate analysis in Nottingham cohort.

	P value	Exp (B)*	95% CI of Exp (B)	
			Lower	Upper
Breast Cancer Specific Survival				
Stage	**$9.8x10^{-8}$**	1.990	1.545	2.563
Grade	**$1.8x10^{-5}$**	1.816	1.383	2.385
HER2 expression	**.001**	1.923	1.304	2.835
BLM protein (Nuclear)	**.026**	.684	.489	.955
BLM protein (Cytoplasmic)	.523	.891	.626	1.269
Rad51 protein (Nuclear)	.156	.797	.583	1.091
Rad51 protein (Cytoplasmic)	.545	1.266	.590	2.717
ER status	.061	1.421	.985	2.050
Lymph node status	.109	1.275	.947	1.715

***B (is a regression coefficient)** - These are the values for the logistic regression equation for predicting the dependent variable from the independent variable. They are in log-odds units. **Exp(B)** - These are the odds ratios for the predictors. They are the exponentiation of the coefficients.

Supplementary Figure Legends

Supplementary Figure S1: A. Forest plot showing prognosis based on BLM protein expression in external validation cohort (n=2413). **B.** Kaplan Meier curves showing metastatic relapse free survival based on BLM protein expression in external validation cohort (n=2413).

Supplementary Figure S2: A1. Western blot of BLM expression in four breast cancer cell lines; MCF-7, MDA-MB-231, MDA-MB-436 and MDA-MB-468. All experiments were run in duplicates. Cells lysates were prepared from 2 million cells and 5 µl of cell lysate was loaded on to the gel ; **A2**. Relative protein expression of BLM in breast Cancer cell lines; **A3**. Relative mRNA expression of BLM in Breast Cancer cell lines. All experiments were run in triplicates**. B1.** Normal breast tissue showing strong nuclear BLM staining. **B2.** Microphotograph of BLM nuclear and BLM cytoplasm negative breast cancer; **B3.** Microphotograph of BLM nuclear negative and BLM cytoplasm positive breast cancer; **B4.** Microphotograph of BLM nuclear positive and BLM cytoplasm negative breast cancer; **B5.** Microphotograph of BLM nuclear positive and BLM cytoplasm positive breast cancer.

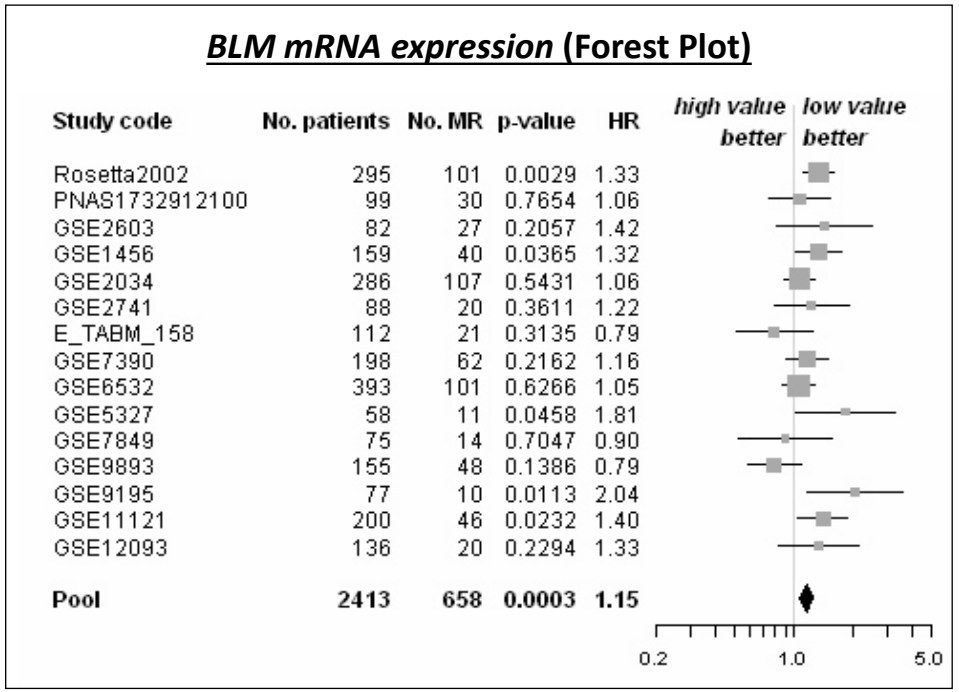
Supplementary Figure S3: Kaplan Meier curves showing BCSS based on BLM protein expression in **A.** ER+ patients with NPI >3.4, who received no endocrine therapy; **B.** ER+ patients with NPI >3.4, who received endocrine therapy; **C.** ER- patients with NPI>3.4, who received no chemotherapy; **D.** ER- patients with NPI >3.4, who received chemotherapy.

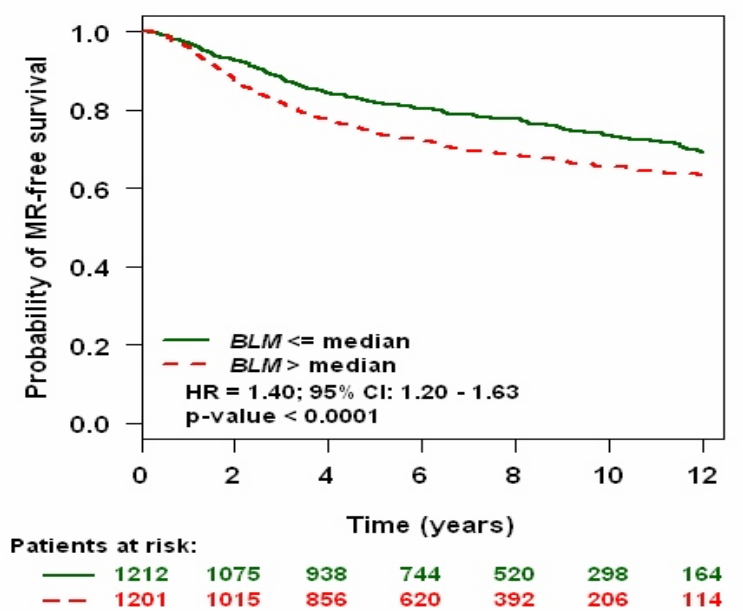
Supplementary Figure S4: Kaplan Meier curves showing BCSS based on BLM/Rad51 protein co-expression in **A.** whole cohort; **B.** ER- cohort with NPI >3.4; **C.** ER- patients with NPI >3.4, who received no chemotherapy; **D.** ER- patients with NPI >3.4, who received chemotherapy.

Supplementary Figure S5: Kaplan Meier curves showing BCSS based on BLM/Rad51 protein co-expression in **A.** ER+ cohort; **B.** ER+ patients with NPI >3.4, who received no endocrine therapy; **C.** ER+ patients with NPI >3.4, who received endocrine therapy.

A

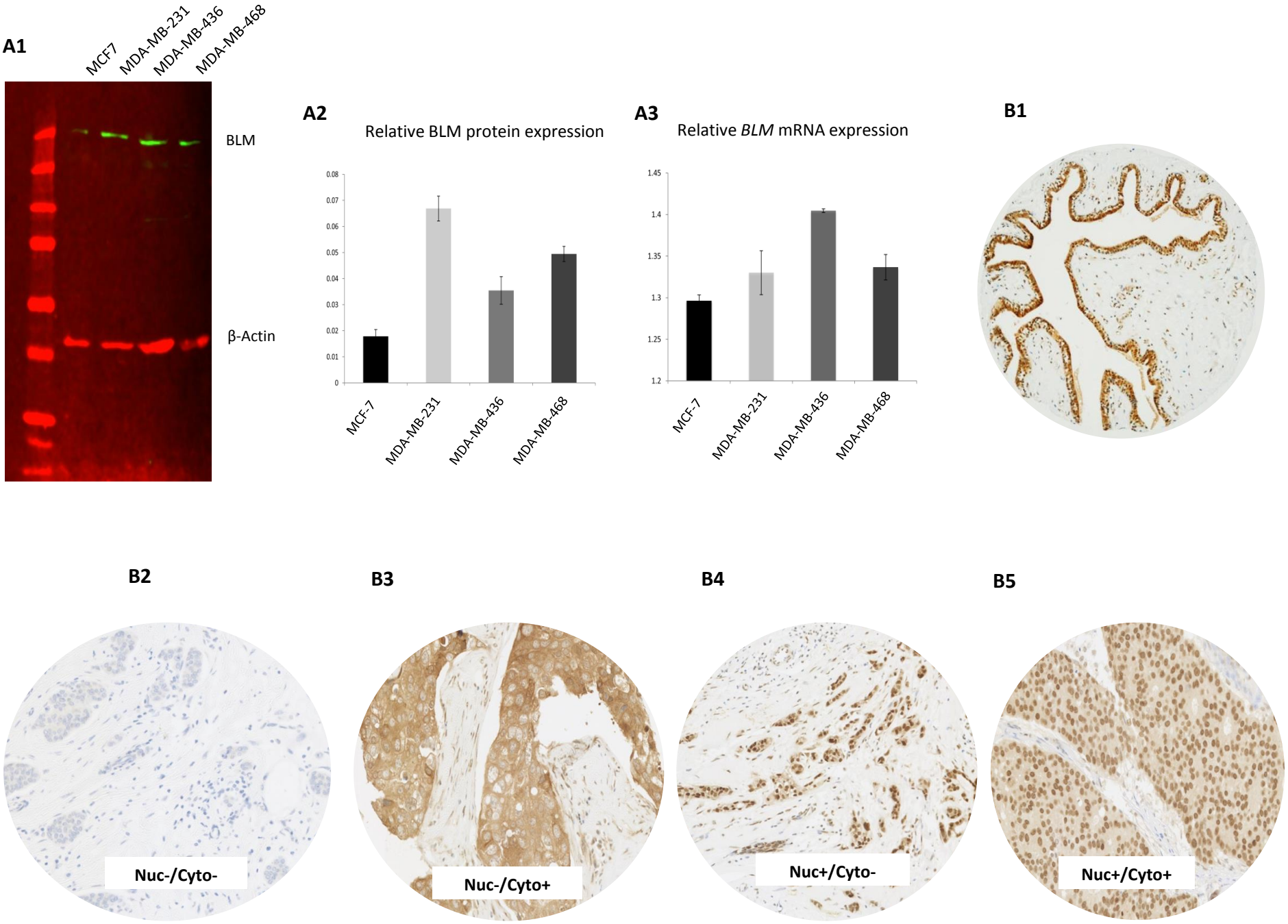
B





MR: Metastatic Relapse

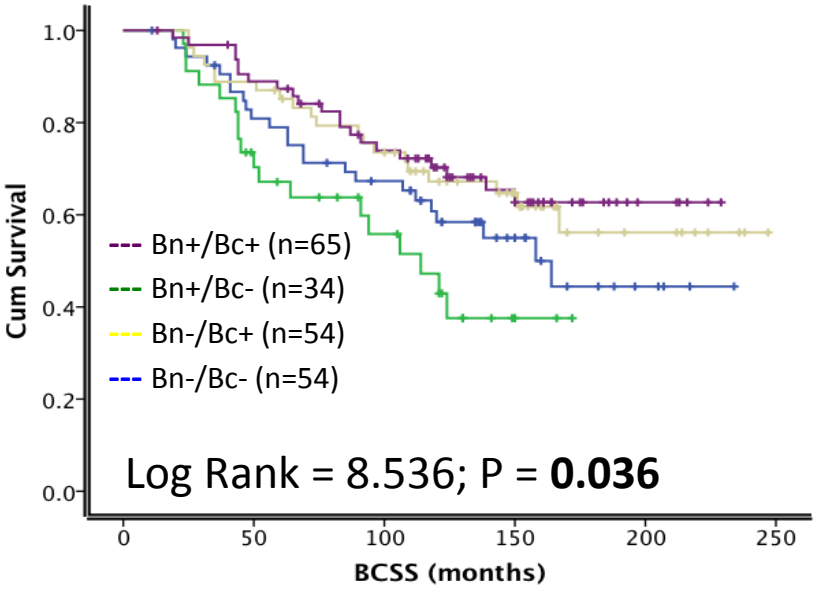
Supplementary Figure S1



Supplementary Figure S2

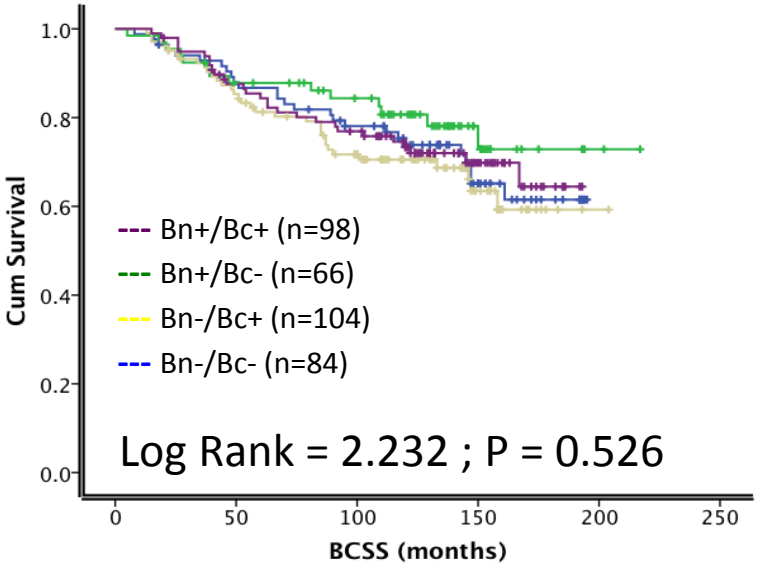
A

ER+; NPI>3.4; no endocrine therapy



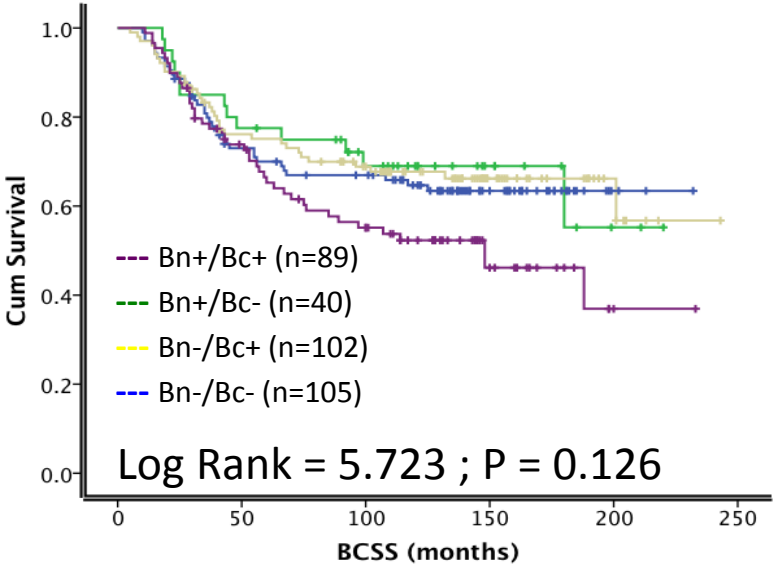
B

ER+; NPI>3.4; had endocrine therapy



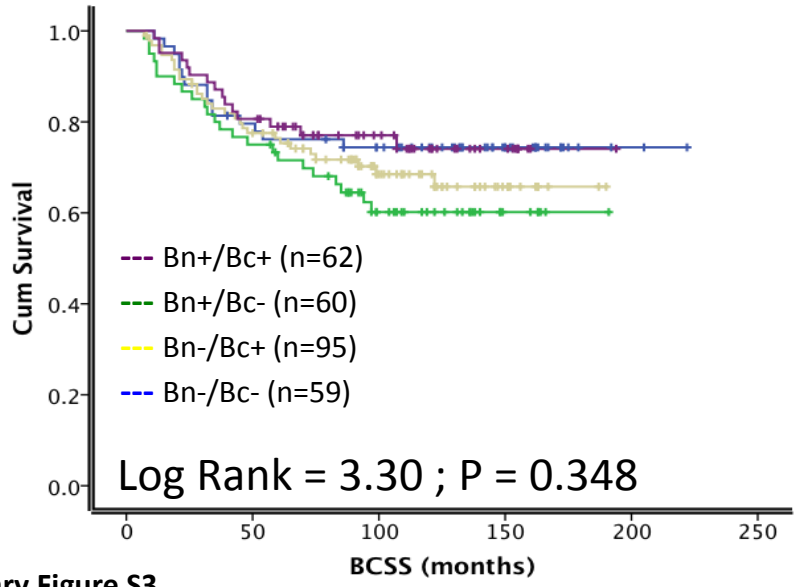
C

ER-; NPI>3.4; no chemotherapy



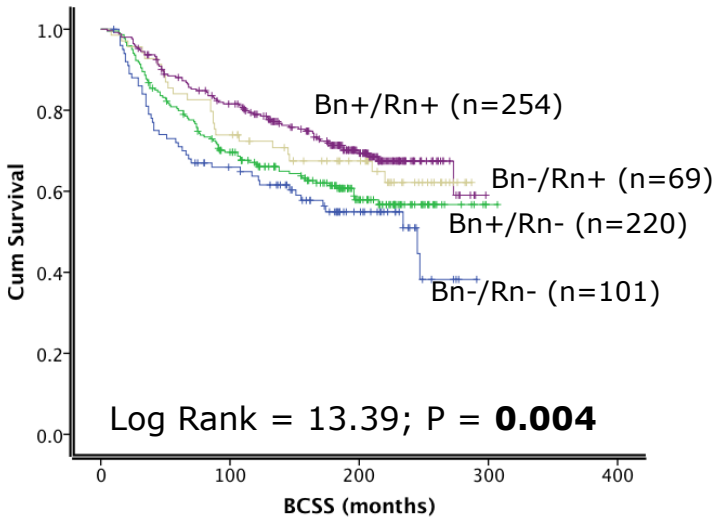
D

ER-; NPI>3.4; had chemotherapy

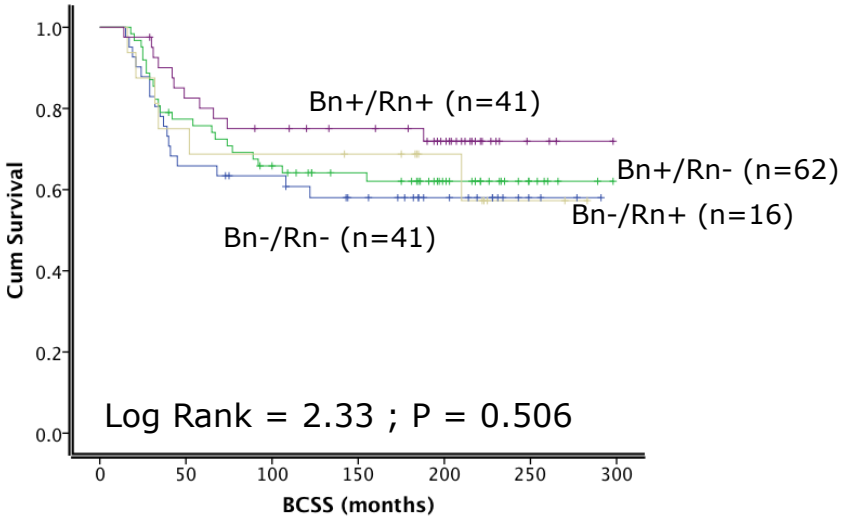


Supplementary Figure S3

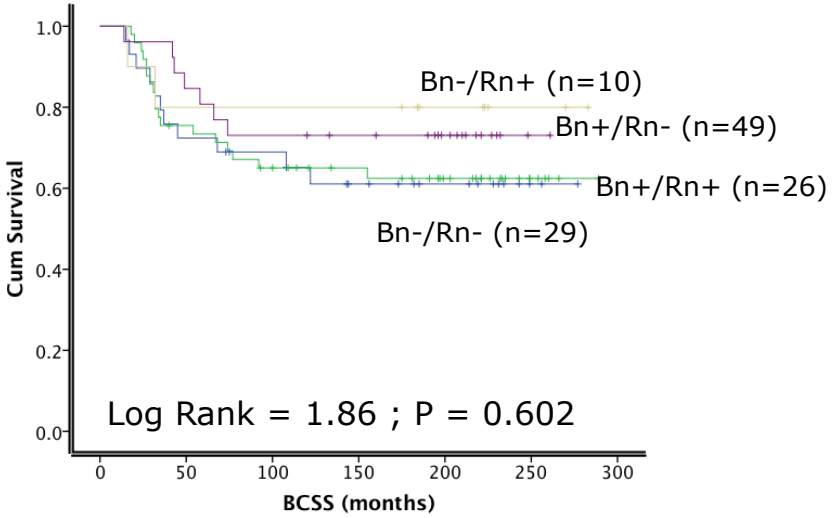
A **BLM Nu/Rad51 co-expression in Whole Cohort**



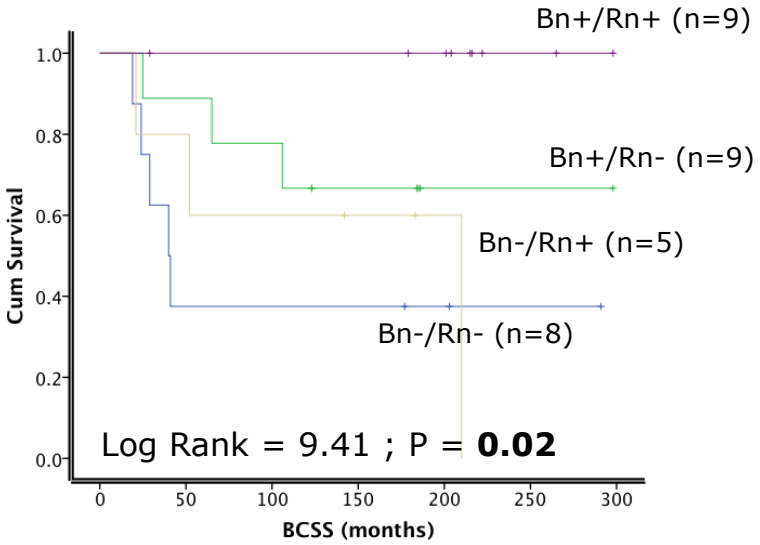
B **NPI>3.4; ER- Cohort**



C **NPI>3.4; ER- Cohort; no chemotherapy**

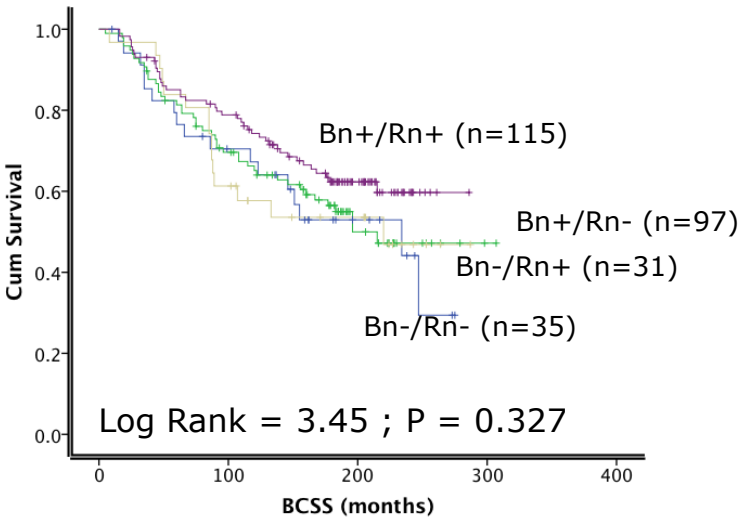


D **NPI>3.4; ER- Cohort; had chemotherapy**

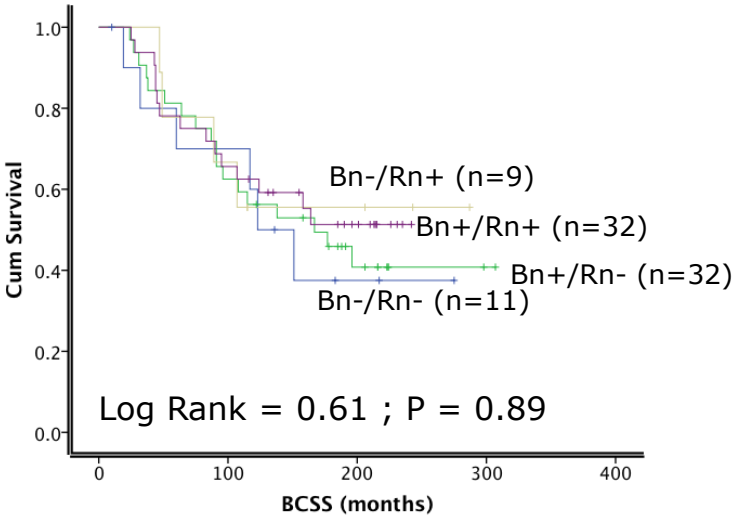


S4Supplementary Figure S4

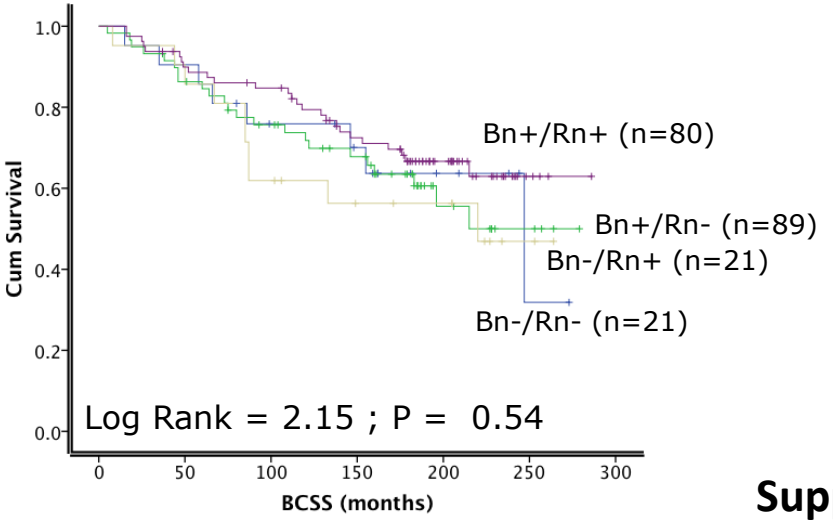
A **BLM Nu/Rad51 co-expression in ER+ Cohort**



B **NPI>3.4; ER+ Cohort; no endocrine therapy**



C **NPI>3.4; ER+ Cohort; had endocrine therapy**



Supplementary Figure S5