Clinicopathological and prognostic significance of RECQL5 helicase expression in breast cancers

Arvind Arora[1,2], Tarek MA Abdel-Fatah[2,] Devika Agarwal[3], Rachel Doherty[1], Deborah L Croteau[4], Paul M. Moseley[2], Khalid Hameed[2], Andrew Green[5], Mohammed A Aleskandarany[5], Emad A Rakha[5], Karl Patterson[6], Graham Ball[3], Stephen YT Chan[2], Ian O Ellis[4], Vilhelm A Bohr[4], Helen E Bryant[6*] and Srinivasan Madhusudan[1,2*]

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

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

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

[4] Laboratory of Molecular Gerontology, Biomedical Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224-6825, USA.

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

[6] Academic Unit of Molecular Oncology, Department of Oncology, Medical School Sheffield Cancer Research Centre, University of Sheffield, Sheffield, S10 2RX.

Running title: RECQL5 in breast cancer

Conflict of interest: The authors disclose no potential conflicts of interest

Word count: 3597

Tables: 4

Figures: 2

* Joint corresponding authors:

Dr Srinivasan Madhusudan, University of Nottingham

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

&

Dr Helen E Bryant, University of Sheffield

E-Mail: h.bryant@sheffield.ac.uk

ABSTRACT

RECQL5 is a member of the RecQ family of DNA helicases and has key roles in homologous recombination, base excision repair, replication and transcription. The clinicopathological significance of RECQL5 expression in breast cancer is unknown. In the current study we have evaluated *RECQL5* mRNA expression in 1977 breast cancers, and RECQL5 protein level in 1902 breast cancers [Nottingham Tenovus series (n=1650) and ER- cohort (n=252)]. Expression levels were correlated to aggressive phenotypes and survival outcomes. High *RECQL5* mRNA expression was significantly associated with high histological grade (p=0.007), HER2 overexpression (p=0.032), ER+/HER2-/high proliferation genefu subtype (p <0.0001), integrative molecular clusters (intClust 1and 9) (p <0.0001) and poor survival (p <0.0001). In sub-group analysis, high *RECQL5* mRNA level remains significantly associated with poor BCSS in ER+ cohort (p<0.0001) but not in ER- cohort (p=0.116). At the protein level, in tumours with low RAD51, high RECQL5 level was significantly associated with high histological grade (p<0.0001), higher mitotic index (p=0.008), de-differentiation (p= 0.025), pleomorphism (p= 0.027) and poor survival (P=0.003). In sub-group analysis, high RECQL5/low RAD51 remains significantly associated with poor BCSS in ER+ cohort (p=0.010), but not in ER- cohort (p=0.628). In multivariate analysis, high *RECQL5* mRNA and high RECQL5/low RAD51 nuclear protein co-expression independently influenced survival (p = 0.022) in whole cohort and in the ER+ sub-group. Pre-clinically, we show that exogenous expression of RECQL5 in MCF10A cells can drive proliferation supporting an oncogenic function for RECQL5 in breast cancer. We conclude that RECQL5 is a promising biomarker in breast cancer.

Summary: RECQL5 is a member of the RecQ family of DNA helicases and has key roles in homologous recombination, base excision repair, replication and transcription. We provide the first clinical evidence that RECQL5 expression is associated with aggressive breast cancers.

INTRODUCTION

RecQ helicases are a highly conserved family of proteins with critical roles in the maintenance of genomic stability (1-4). RECQL5 is a key member of the mammalian RecQ helicase family (1). It is a 3'-5' helicase with weak Holiday junction unwinding activity (1). RECQL5 preferentially unwinds the lagging strand of DNA at replication fork structures and can promote strand exchange *in vitro* (5). Three isoforms of RECQL5 have been identified; RECQL5α (410aa), RECQL5β (435aa) and RECQL5γ (991aa) (6). Whereas RECQL5α and RECQL5γ are cytoplasmic, RECQL5β isoform has nuclear localisation (6). RECQL5β (here in referred to as RECQL5) interacts with multiple DNA repair and metabolising proteins including RAD51, RAD50, PARP1, FEN1, RNA Polymerase II, WRN, BLM and Mre11, NBS1, PCNA, TopIIa and Top IIIa/b (1). RECQL5 has important roles in homologous recombination (HR), base excision repair (BER), DNA replication and transcription (1). Evidence for a role in HR includes findings that RECQL5 deficient cells have an increased rate of sister chromatid exchange (7), and that depleted cells accumulate γH2Ax and RAD51 foci (8). Mechanistically RECQL5's function in HR can be proposed from *in vitro* studies showing that RECQL5 physically interacts with RAD51, causing an ATPase dependent disruption of RAD51 mediated presynaptic filament formation and hence has anti-recombinase activity (9) similar to Bloom syndrome helicase (BLM) (7). Further RECQL5 associates with the MRN complex and inhibits MRE11 exonuclease activity (10), which may also contribute to the HR function of RECQL5. Interestingly while the MRN complex is required to bring RECQL5 to sites of DSBs, it's recruitment is independent of the helicase activity (11). RECQL5 may also have a role in base excision repair (BER). It's precise function in this pathway is not clear but RECQL5 interacts with the long patch BER proteins

PCNA and FEN1, it's interaction stimulating FEN1 activity, and RECQL5 co-localises with FEN1 after oxidative damage of DNA (12). In addition the short patch BER protein XRCC1 is retained at sites of oxidative damage in the absence of RECQL5 and expression of PARP1 and XRCC1 maybe regulated by RECQL5 (13). A role in protecting cells from DNA replication stress has also been shown for RECQL5. RECQL5 foci increase following replication stress and overexpressing RECQL5 can overcome thymidine induced replication stress (14). Likewise RECQL5 can prevent spontaneous replication fork collapse and RECQL5 depleted cells are hypersensitive the DNA replication inhibitor camptothecin, (15, 16). RECQL5 has also been implicated in regulation of transcription elongation and can supress genomic instability associated with transcriptional stress (2, 17, 18).

Germline mutation of three of the RecQ helicases (BLM, WRN and RECQL4) leads to cancer predisposition syndromes namely Bloom syndrome, Werner syndrome and Rothmund-Thomson syndrome (3, 4). Although RECQL5 helicase has not been associated with any disease phenotype, in preclinical studies, RECQL5 deficient mice cells show increased levels of spontaneous double strand breaks, are susceptible to gross chromosomal rearrangements and are prone to develop lymphomas and various solid tumours including breast cancer (8) (19).

We hypothesised that RECQL5 may influence breast cancer pathogenesis. In the current study we have evaluated *RECQL5* mRNA expression in 1950 breast cancers, and RECQL5 protein level in 1902 breast cancers. We provide the first clinical evidence that RECQL5 may influence the development of aggressive breast cancer and have prognostic significance

particularly in ER+ breast cancers. In addition we show that exogenous expression of RECQL5 in MCF10A cells can drive proliferation supporting a oncogenic function for RECQL5 in breast cancer.

MATERIALS AND METHODS

RECQL5 and mRNA expression: *RECQL5* mRNA expression was investigated in METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohort. The METABRIC study protocol, detailing the molecular profiling methodology in a cohort of 1980 breast cancer samples is described by Curtis et al (20). Patient demographics are summarized in **supplementary Table S1** of supporting information. ER positive and/or lymph node negative patients did not receive adjuvant chemotherapy. ER negative and/or lymph node positive patients received adjuvant chemotherapy. For this cohort, the mRNA expression was hybridized to Illumina HT-12 v3 platform (Bead Arrays), and the data were pre-processed and normalized as described previously (20). RECQL5 expression was evaluated in this data set (RECQL5 probe ID: ILMN_1697682). The probe was a perfect match and quality for its target, having a GC content of 58% , 0 SNPs and it does not possess a polyG tail at the end. Samples were classified into the intrinsic subtypes based on the PAM50 gene list. A description of the normalization, segmentation, and statistical analyses was previously described (20). Real time RT-qPCR was performed on the ABI Prism 7900HT sequence detection system (Applied Biosystems) using SYBR1 Green reporter. All the samples were analysed as triplicates. The Chi-square test was used for testing association between categorical variables, and a multivariate Cox model was fitted to the data using as endpoint breast cancer specific death. X-tile (Version 3.6.1) was used to identify a cut-off in gene expression values such that the resulting subgroups have

RECQL5 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 S2**. This is a well-characterized series of patients with long-term follow-up that have been investigated in a wide range of biomarker studies (22-25). All patients were treated in a uniform way in a single institution with standard surgery (mastectomy or wide local excision), followed by 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), oestrogen receptor-α (ER-α) 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-α positive tumours were also

offered endocrine therapy. Postmenopausal patients with NPI scores of ≥3.4 and ER-α

positivity were offered endocrine therapy, while ER-α negative patients received classical

CMF chemotherapy. Median follow up was 111 months (range 1 to 233 months). Survival

data, including overall survival, disease-free survival (DFS), and development of loco-

regional and distant metastases (DM), was maintained on a prospective basis. DFS was

defined as the number of months from diagnosis to the occurrence of local recurrence, local

lymph node (LN) relapse or DM relapse. Breast cancer specific survival (BCSS) was defined

as the number of months from diagnosis to the occurrence of BC related-death. Local

recurrence free survival (LRS) was defined as the number of months from diagnosis to the

occurrence of local recurrence. DM-free survival was defined as the number of months from

diagnosis to the occurrence of DM relapse. 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 an independent series of 252 ER-α negative invasive BCs diagnosed and

managed at the Nottingham University Hospitals between 1999 and 2007. All patients were

primarily treated with surgery, followed by Radiotherapy and anthracycline chemotherapy.

The characteristics of this cohort are summarised in **supplementary Table S3**.

Tumor Marker Prognostic Studies (REMARK) criteria, recommended by McShane et al (26),

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 RECQL5 and

other biological antibodies (**Supplementary Table S4** of supporting information) as

previously described (22-25). 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). Leica Autostainer XL machine was used to dewax and rehydrate the slides. Pre-treatment antigen retrieval was performed on the TMA sections using sodium citrate buffer (pH 6.0) and heated for 20 minutes at 95^0C in a microwave (Whirpool JT359 Jet Chef 1000W). A set of slides were incubated for 60 minutes with the primary anti-RECQL5 antibody (HPA029971, Sigma-Aldrich, UK) and anti-RAD51 antibody (clone Ab88572, Abcam Ltd., Cambridge, UK), at a dilution of 1:100 and 1:70 respectively. 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 (AA and TAF) 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%). H-score (range 0-300) was calculated by multiplying intensity of staining and percentage staining. RECQL5, RAD51 and TOPO2A expression was categorised based on the frequency histogram distributions. A median H score of ≥10 was taken as the cut-off for high RECQL5 and a median H score of ≥8 was taken as the cut-off for high RAD51 nuclear expression. For TOPO2A >25% staining cells were taken

as high TOPO2A 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 Holm-Bonferroni correction method (27).

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) authenticated cell lines were purchased from ATCC and 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-RECQL5 antibody (HPA029971, Sigma-Aldrich,UK) was incubated over night at room temperature at a dilution of 1:2000. 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.

RECQL5 knockdown breast cancer cells using siRNAs: MCF7 cells were transfected with 50 nM RECQL5 siRNA (Dharmacon, Lafayette, CO) using *DharmaFECT* 1 reagent (Dharmacon, Lafayette, CO) and left for 48 hours. The cell lysates were prepared and western blotted for RECQL5 as above but immunoreactive protein was visualized using ECL reagents (Amersham Pharmacia) following manufacturer's instructions. The sequences used for Q5β-1 and Q5β-3 siRNA constructs were 5'-UGAAGAAGGUGGCCGAUAU-3' and 5'-CUGCAAAUGUUGUGGUCAA-3' respectively.

RECQL5 overexpression in MCF10A cells and proliferation assay: 10,000 MCF10A cells were plated into replica wells of 6 well plate and transfected with pcDNA-RECQL5β (a generous gift from Dr. Pavel Janscak - Institute of Molecular Cancer Research, University of Zurich) or empty vector control. Following trypsinisation to remove cells from the plate live cells counted every 24 h post transfection using trypan blue exclusion. In parallel cell lysates were prepared and western blotted for RECQL5 as above but immunoreactive protein was visualized using ECL reagents (Amersham Pharmacia) following manufacturer's instructions.

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 (RECQL5 QuantiTect Prier Assay,Cat. No.

QT00084973, QIAGEN) targeting RECQL5 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 *RECQL5* mRNA levels associate with aggressive sporadic breast cancer: We initially profiled a panel of breast cancer cell lines. MDA-MB-436, MDA-MB-468 MCF-7 and MDA-MB 231 have robust expression of *RECQL5* mRNA (**Figure 1A**). We then proceeded to investigate *RECQL5* mRNA expression in the METABRIC cohort. 1306/1977 (66%) of breast tumours had low *RECQL5* mRNA expression and 34% (671/1977) of breast tumours had high *RECQL5* mRNA expression. High *RECQL5* mRNA expression was significantly associated with aggressive clinicopathological features including high histological grade, HER2 over expression, and triple negative phenotypes (ps<0.05) **[Table 1]**. High *RECQL4* mRNA expression was also found to be significantly correlated to previously described breast cancer molecular phenotypes; PAM50.Her2 (p<0.0001) and Genufu subtype (ER+/Her2-/High proliferation) (p<0.01) breast tumours **[Table 1].** On the other hand, PAM50.LumA and Genufu subtypes (ER+/Her2-/low proliferation) were associated with low levels of *RECQL5* mRNA (ps<0.05) **[Table 1]**. Interestingly, PAM50.Basal was also more common in tumours with low levels of *RECQL5* mRNA (ps<0.05). Similarly, low *RECQL5* mRNA expression was associated with intClust.3 (p<0.00001) and intClust.4 (ps<0.001) molecular phenotypes that have good prognosis (28). However, high *RECQL5* mRNA expression was significantly associated with intClust.1 (p<0.00001), and intClust.9 (p<0.00001) that have intermediate to poor clinical outcome (20).

We then proceeded to survival analysis in METABRIC cohort. High *RECQL5* mRNA expression was associated with poor breast cancer specific survival (BCSS) (p<0.0001) in the

whole cohort **[Figure 1B]**. In the ER+ sub-group, high *RECQL5* mRNA expression was found to be associated with poor BCSS (p<0.0001) **[Figure 1C]**. In the ER+ sub-group that received adjuvant endocrine therapy, high *RECQL5* mRNA expression remained associated with poor BCSS (p<0.001) **[Figure 1D]**. However, in ER- sub-group, *RECQL5* mRNA expression did not significantly influence outcome in the ER- cohort, including ER- patients who received chemotherapy (p=0.116 and p=0.213 respectively) **[Figure 1E, 1F]**.

In the multivariate Cox regression analysis, that included other validated prognostic factors, *RECQL5* mRNA expression was a powerful independent predictor for breast cancer specific survival in the whole cohort (p=0.038), ER+ sub-group (p=0.046) but not in the ER- sub-group (p=0.615) (**Table 2**).

Together the data provides evidence that *RECQL5* mRNA level has clinicopathological significance and influence prognosis, particularly in ER+ breast cancers. We proceeded to evaluate RECQL5 protein expression in breast cancers.

RECQL5 protein expression in sporadic breast cancer: We initially investigated RECQL5 protein level in breast cancer cell lines and in MCF10A breast epithelial cells. As shown in **Figure 2A,** MDA-MB-436, MDA-MB-468 MCF-7 and MDA-MB 231 have robust expression of RECQL5 protein. On the other hand, in MCF10A cells, RECQL5 expression was low compared to MCF7 cells (**Figure 2B**). When RECQL5 was overexpressed in MCF10A (**Figure 2B**) we observed increased proliferation in RECQL5 overexpressing MCF10A cells compared to control MCF10A cells (**Figure 2C**). Taken together, the data suggest that RECQL5 is a marker of proliferation and could have prognostic and/or

predictive significance in human breast cancers. We proceeded to IHC investigation in human tumours. To provide evidence for the specificity of the anti-RECQL5 antibody used for IHC studies, we generated RECQL5 knockdown (KD) MCF7 cells using two siRNA constructs. As shown in **Figure 2D**, control cells have robust RECQL5 expression seen as a single band while RECQL5 KD cells show almost complete loss of this band − demonstrating the specificity of the RECQL5 antibody.

We observed nuclear-only localization of RECQL5 protein in breast cancers **[Figure 2D].** There was no cytoplasmic staining in breast tumours. 46.3% (556/1200) of tumours showed low nuclear RECQL5 expression and 53.7% (644/1200) of tumours revealed high nuclear RECQL5 expression. We also investigated RECQL5 expression in 30 normal breast tissues. There was no cytoplasmic staining in normal tissues. We observed that 30/30 (100%) normal tissues showed high nuclear RECQL5 expression. Taken together, the data suggest differential expression RECQL5 in normal tissues compared to tumours tissues. We initially investigated RECQL5 protein alone in breast cancer tissues. Tumours with low nuclear RECQL5 levels were significantly associated with high grade, high mitotic index, tubule formation, tumour type, HER2 overexpression, ER-, triple negative and basal type phenotype (ps<0.05) **[Supplementary Table 5]** . In addition, low RECQL5 expression was found to be associated with low levels of other DNA repair proteins such as BRCA1, XRCC1, FEN1, SMUG1, APE1, Polβ, ATM, ATR, Chk1, Chk2, TOP2A, Rad51, and DNA-PKc (ps< 0.01) **[Supplementary Table 5].** We then conducted the univariate survival analysis. At protein level, RECQL5 expression alone failed to show any statistically significant correlation with breast cancer specific survival in whole cohort **[Supplementary Figure S1A].** We then proceeded to subgroup analysis and again found no statistically significant association with

survival in ER+ **[Supplementary Figure S1B]** and ER- subgroups **[Supplementary Figure S1C]**. As RECQL5 alone, despite having clinicopathological associations with aggressive phenotype, did not have prognostic significance we hypothesised that RECQL5 may operate in the context of RAD51 or TOPO2A to influence clinical outcomes. We therefore proceeded to co-expression analysis.

RECQL5 co-expression with RAD51 or TOPO2A in sporadic breast cancer: RECQL5 physically interacts with RAD51 and disrupts RAD51 mediated presynaptic filament formation (1, 9). Tumours with high RECQL5-low RAD51 nuclear protein levels were significantly associated with high grade, high mitosis and pleomorphism (all p values <0.05) [**Table 3**]. Tumours with low RECQL5-low RAD51 nuclear protein levels were significantly associated with tubule formation, NPI>3.4 and ER/PR negativity (all p values <0.05) [**Table 3**]. In univariate analysis, tumours with high RECQL5-low RAD51 nuclear protein levels were associated with poor breast cancer specific survival in the whole cohort (p=0.003) [**Figure 2E**]. In sub-group analysis, ER+ tumours with high RECQL5-low RAD51 nuclear protein levels were associated with poor breast cancer specific survival (p=0.010) **[Figure 2F]**. In ER- cohort, RECQL5-RAD51 co-expression did not influence survival (p=0.628) [**Supplementary Figure S1D**]. We have previously shown a direct interaction between RECQL5 and Topoisomerase IIα (TOP2A) where RECQL5 specifically stimulated the decatenation activity of TOP2A (29). Tumours with low RECQL5-high TOPO2A were significantly associated with high grade, high mitotic index, de-differentiation tumour type, high risk NPI and PR negativity (**Supplementary Table S6**). In univariate analysis, RECQL5-TOPO2A co-expression did not influence survival (**Supplementary Figures S1E, S1F, S1G**)

In multivariate analysis (**Table 4**), RECQL5-RAD51 co-expression remained statistically significant independent marker of prognosis (p=0.022) in the whole cohort. NPI and HER-2 expression were other factors independently associated with breast cancer specific survival (p values 0.001 and 0.0001 respectively).   In the ER+ sub-group, RECQL5-RAD51 co-expression was of borderline significance (p=0.07), but not in the ER- sub-group (p=0.172).

DISCUSSION

RECQL5 is a key member of the RecQ family of DNA helicases (1-4). RECQL5 is a multifunctional protein with roles in DNA replication, chromosomal segregation, DNA repair (including homologous recombination, single strand break repair and base excision repair) and transcription. Deficiency of DNA repair factors such as RECQL5 could predispose an individual to cancer through increased genomic instability. On the other hand, in established tumours high levels of RECQL5 may promote survival of cancer cells that are constantly under DNA damaging oxidative stress. RECQL5 deficient mice are predisposed to lymphomas and solid tumours including breast tumours (1-4). In humans, polymorphisms within the *RECQL5* gene may predispose an individual to cancer including: osteosarcomas, laryngeal carcinomas and breast cancers (30-33). In a recent study in colorectal cancers, low RECQL5 expression was observed at the mRNA and protein levels (19). The authors concluded that RECQL5 deficiency may predispose to colorectal cancer. However, the study had several limitations, including a small cohort and lack of clear evidence of clinicopathological associations or survival outcomes (19).

The clinicopathological significance of RECQL5 in sporadic breast cancer is unknown. In the current study we have comprehensively evaluated RECQL5, and unravelled its complex role in breast cancers. At the mRNA level, we observed high expression in 34% of tumours. High *RECQL5* mRNA expression was significantly associated with aggressive phenotypes and adverse survival. The prognostic significance was particularly pronounced in ER+ sub-group, supporting the observation that high *RECQL5* mRNA expression was more likely in ER+/ HER2 -/high proliferation Geneufu sub-type tumours. At the protein level, we observed high

20

RECQL5 levels in 53.7% of breast cancers. However it should be pointed out that RECQL5 mRNA expression and RECQL5 protein expression were investigated in two independent cohorts. Therefore a limitation to the current study is that we were unable to perform direct correlations between mRNA and protein expression. In contrast to mRNA expression data, although low RECQL5 expression was associated with aggressive phenotype, RECQL5 protein alone did not influence survival outcomes in patients. The data suggest that RECQL5 may be subjected to complex post-transcriptional regulation. Alternatively, there may be a threshold effect for expression level and this threshold is more accurately measured by the analysis of mRNA levels. Another possibility is that the complex role of RECQL5 may operate in the context of RAD51. RECQL5 directly interacts with RAD51 (9) and the anti-recombinogenic role of RECQL5 may be active through disruption of RAD51 mediated presynaptic filament formation during HR (34). Therefore, we conducted RECQL5-RAD51 co-expression studies and observed that tumours with high RECQL5/low RAD51 not only manifest aggressive phenotypes but were also associated with poor survival. In sub-group studies, similar to mRNA data, the prognostic significance was more pronounced in ER+ breast cancer but not in ER- tumours. Consistent with high RECQL5 being associated with aggressive phenotypes, over expression of RECQL5 in the normal breast cell line MCF10A increased proliferation. Interestingly RAD51 is reported as having relatively low expression in MCF10A cells (35)and the functional relationship between RECQL5, RAD51, proliferation, recombination and tumourigenesis will be the subject of future investigations. Taken together, the data provides the first clinical evidence that RECQL5 may influence breast cancer pathogenesis.

RECQL5 interacts with and is a general transcription elongation factor for RNA Pol II (2, 17, 18). Loss of RECQL5 leads to a genome-wide increase in the average rate of gene transcription, transcriptional stress and recombination (17) suggesting RECQL5 has a function in resolving such stress. RNA polymerases generate positive supercoiling ahead of the transcription apparatus, which in turn feeds back to reduce the processional rate of RNA polymerase. Topoisomerases are ubiquitous enzymes that remove supercoiling and therefore are required for normal transcription elongation. Specifically, type II topoisomerases generate transient protein-concealed double strand breaks while removing torsional stress from the DNA and collisions with the transcription apparatus can convert these protein-DNA complexes into permanent DNA strand breaks (36). Previously, we showed a direct interaction between RECQL5 and Topoisomerase IIα (TOP2A) specifically stimulated the decatenation activity of TOP2A (29). TOP2A expression is a marker for proliferation and has been analyzed in a number of breast cancer studies in part because it lies in close proximity to the HER2 gene on chromosome 17. Therefore, we were interested in evaluating if there were any expression correlations between RECQL5, which also resides on chromosome 17 (17q25.1), and TOP2A expression in our breast cancer cohorts. Interestingly, high grade tumours are more often scored as RECQL5- and TOP2A+, than any other expression pattern. This may reflect the fact that loss of RECQL5 promotes transcriptional stress. Additionally, while TOP2A is known as a marker for proliferation, loss of RECQL5 may make these cells more dependent on TOP2A to relieve transcriptional stress. Thus in this subgroup of patients, it would be interesting to evaluate the efficacy of TOP2A targeted therapies (such as doxorubicin that is routinely used in breast cancer therapy), if in fact these tumours are more dependent on topoisomerase activities than other tumour types cells.

In conclusion, our data provides evidence that RECQL5 could be a promising biomarker in breast cancer and needs further investigation as a potential drug target.

Competing financial interests: The authors declare no competing financial interests

REFERENCES

1.   Popuri, V., et al. (2013) Human RECQL5: guarding the crossroads of DNA replication and transcription and providing backup capability. Crit Rev Biochem Mol Biol, **48**, 289-99.
2.   Aygun, O., et al. (2010) RECQL5 helicase: connections to DNA recombination and RNA polymerase II transcription. DNA Repair (Amst), **9**, 345-53.
3.   Brosh, R.M., Jr. (2013) DNA helicases involved in DNA repair and their roles in cancer. Nat Rev Cancer, **13**, 542-58.
4.   Hanada, K., et al. (2007) Molecular genetics of RecQ helicase disorders. Cell Mol Life Sci, **64**, 2306-22.
5.   Kanagaraj, R., et al. (2006) Human RECQ5beta helicase promotes strand exchange on synthetic DNA structures resembling a stalled replication fork. Nucleic Acids Res, **34**, 5217-31.
6.   Shimamoto, A., et al. (2000) Human RecQ5beta, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3alpha and 3beta. Nucleic Acids Res, **28**, 1647-55.
7.   Hu, Y., et al. (2005) Recql5 and Blm RecQ DNA helicases have nonredundant roles in suppressing crossovers. Mol Cell Biol, **25**, 3431-42.
8.   Hu, Y., et al. (2007) RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. Genes Dev, **21**, 3073-84.
9.   Schwendener, S., et al. (2010) Physical interaction of RECQ5 helicase with RAD51 facilitates its anti-recombinase activity. J Biol Chem, **285**, 15739-45.
10.   Zheng, L., et al. (2009) MRE11 complex links RECQ5 helicase to sites of DNA damage. Nucleic Acids Res, **37**, 2645-57.
11.   Popuri, V., et al. (2012) Recruitment and retention dynamics of RECQL5 at DNA double strand break sites. DNA Repair (Amst), **11**, 624-35.
12.   Speina, E., et al. (2010) Human RECQL5beta stimulates flap endonuclease 1. Nucleic Acids Res, **38**, 2904-16.
13.   Tadokoro, T., et al. (2012) Human RECQL5 participates in the removal of endogenous DNA damage. Mol Biol Cell, **23**, 4273-85.
14.   Blundred, R., et al. (2010) Human RECQL5 overcomes thymidine-induced replication stress. DNA Repair (Amst), **9**, 964-75.
15.   Hu, Y., et al. (2009) Recql5 plays an important role in DNA replication and cell survival after camptothecin treatment. Mol Biol Cell, **20**, 114-23.
16.   Wang, X., et al. (2011) RECQL5 is an important determinant for camptothecin tolerance in human colorectal cancer cells. Biosci Rep, **31**, 363-9.
17.   Saponaro, M., et al. (2014) RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell, **157**, 1037-49.
18.   Izumikawa, K., et al. (2008) Association of human DNA helicase RecQ5beta with RNA polymerase II and its possible role in transcription. Biochem J, **413**, 505-16.
19.   Lao, V.V., et al. (2013) Altered RECQ Helicase Expression in Sporadic Primary Colorectal Cancers. Transl Oncol, **6**, 458-69.
20.   Curtis, C., et al. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, **486**, 346-52.

21.   Camp, R.L., et al. (2004) X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res, **10**, 7252-9.

22.   Arora, A., et al. (2015) Transcriptomic and Protein Expression Analysis Reveals Clinicopathological Significance of Bloom Syndrome Helicase (BLM) in Breast Cancer. Mol Cancer Ther, **14**, 1057-65.

23.   Abdel-Fatah, T.M., et al. (2015) Untangling the ATR-CHEK1 network for prognostication, prediction and therapeutic target validation in breast cancer. Mol Oncol, **9**, 569-85.

24.   Abdel-Fatah, T.M., et al. (2014) Clinicopathological significance of ATM-Chk2 expression in sporadic breast cancers: a comprehensive analysis in large cohorts. Neoplasia, **16**, 982-91.

25.   Abdel-Fatah, T.M., et al. (2014) Is there a role for base excision repair in estrogen/estrogen receptor-driven breast cancers? Antioxid Redox Signal, **21**, 2262-8.

26.   McShane, L.M., et al. (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst, **97**, 1180-4.

27.   Holm, S. (1979) A simple sequentially rejective multiple test procedure. Scand J Stat, **6**, 65-70.

28.   Curtis, C., et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, **486**, 346-52.

29.   Ramamoorthy, M., et al. (2012) RECQL5 cooperates with Topoisomerase II alpha in DNA decatenation and cell cycle progression. Nucleic Acids Res, **40**, 1621-35.

30.   Dong, Y.Z., et al. (2015) Single nucleotide polymorphism in the RECQL5 gene increased osteosarcoma susceptibility in a Chinese Han population. Genet Mol Res, **14**, 1899-902.

31.   He, Y.J., et al. (2014) Association between RECQL5 genetic polymorphisms and susceptibility to breast cancer. Tumour Biol, **35**, 12201-4.

32.   Zhi, L.Q., et al. (2014) Association of RECQL5 gene polymorphisms and osteosarcoma in a Chinese Han population. Tumour Biol, **35**, 3255-9.

33.   Qi, Y., et al. (2014) Haplotype analysis of RECQL5 gene and laryngeal cancer. Tumour Biol, **35**, 2669-73.

34.   Islam, M.N., et al. (2012) A variant of the breast cancer type 2 susceptibility protein (BRC) repeat is essential for the RECQL5 helicase to interact with RAD51 recombinase for genome stabilization. J Biol Chem, **287**, 23808-18.

35.   Yuan, S.S., et al. (1999) BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo. Cancer Res, **59**, 3547-51.

36.   Ashour, M.E., et al. (2015) Topoisomerase-mediated chromosomal break repair: an emerging player in many games. Nat Rev Cancer, **15**, 137-51.

Table 1: Association between *RecQL5* mRNA expression and clinico-pathologic variables in METABRIC cohort.

VARIABLE	RECQL5 mRNA Expression		P Values	
	Low	**High**		
	N(%)	N (%)	*Unadjusted*	*Adjusted*
A) Pathological Parameters				
Lymph node stage				
Negative	684(52.6%)	351(52.5%)	0.990	25.74
Positive (1-3)	207(15.9%)	107(15.9%)		
Positive (>3)	409(31.5%)	213(31.7%)		
Grade				
G1	129(10.3%)	40(6.2%)	**0.007**	**0.023**
G2	512(41.0%)	258(40.2%)		
G3	607(48.6%)	343(53.5%)		
Tumour Size (cm)				
T 1a+b(1.0)	61(4.7%)	31(4.7%)	0.097	0.157
T 1c(>1.0-2.0)	532(41.0%)	234(35.6%)		
T2 (>2.0-5)	639(49.2%)	362(55.0%)		
T3 (>5)	67(68.4%)	31(31.6%)		
NPI				
≤ 3.4	461(35.3%)	219(32.6%)	0.238	0.269
>3.4	137(64.7%)	1160(67.4%)		
Her2 overexpression (No)	1159(88.7%)	573(85.4%)	**0.032**	0.075
(Yes)	147(11.3%)	98(14.6%)		
Triple negative (No)	1079(82.6)	581 (88.6)	**0.023**	0.066
(Yes)	227(17.4)	90(13.4)		
ER (Negative)	322(24.7%)	148(22.1%)	0.199	0.246
(Positive)	984(75.3%)	523(77.9%)		
PgR (Negative)	603(46.2%)	333(49.6%)	0.145	0.188
(Positive)	703(53.8%)	338(50.4%)		

Genefu subtype				
ER-/Her-2 negative	110(16.3%)	40(12.6%)	0.129	0.186
ER+/Her-2 negative/high proliferation	229(33.9%)	137(43.1%)	**0.005**	**0.018**
ER+/Her-2 negative/low proliferation	265(39.1%)	103(32.4%)	**0.038**	**0.082**
Her-2 positive	72(10.7%)	38(11.9%)	0.543	0.564
PAM50 subtype				
PAM50.Her2	130(11.2%)	108(17.6%)	**1.5×10^{-4}**	**0.0008**
PAM50.Basal	233(20.1%)	97(15.8%)	**0.029**	0.075
PAM50.LumA	488(42.1%)	227(37.1%)	**0.042**	0.084
PAM50.LumB	309(26.6%)	180(29.4%)	0.214	0.252
IntClust subgroups				
intClust.1	49(3.8%)	88(13.1%)	**8.3×10^{-15}**	**0.00001**
intClust.2	55(4.2%)	17(2.5%)	0.059	0.102
intClust.3	221(16.9%)	69(10.3%)	**7.8×10^{-5}**	0.0005
intClust.4	267(20.4%)	76(11.3%)	**3.9×10^{-7}**	**0.00001**
intClust.5	120(63.5%)	69(36.5%)	0.433	0.469
intClust.6	50(3.8%)	36(5.4%)	0.113	0.172
intClust.7	113(8.7%)	40.2(11.3%)	0.056	0.104
intClust.8	187(14.3)	113(16.8%)	0.139	0.191
intClust.9	74(5.7%)	72(10.7%)	**4.6×10^{-5}**	**0.0004**
intClust.10	170(13.0%)	55(8.2%)	**0.001**	**0.0043**

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.

Table 2. Multivariate analysis of RECQL5 mRNA expression in breast cancer

	P-Value	Risk Ratio	95% CI for Risk Ratio	
			Lower	Upper
Breast Cancer Specific Survival (Whole Cohort)				
Grade	**0.000004**	1.447429	1.237421	1.693079
Size (Cm)	**0.000001**	1.121003	1.076755	1.167069
LN Status	**0.000001**	1.875371	1.666800	2.110040
RECQL5 mRNA expression	**0.023530**	1.223718	1.027549	1.457337
Breast Cancer Specific Survival (ER+ Cohort)				
Grade	**0.0002**	1.411263	1.176981	1.692178
Size (Cm)	**0.000001**	1.180525	1.123389	1.240566
LN Status	**0.000001**	1.822326	1.578493	2.103825
RECQL5 mRNA expression	**0.045**	1.242008	1.004707	1.535359
Breast Cancer Specific Survival (ER- Cohort)				
Grade	0.968564	1.008054	0.676448	1.502217
Size (Cm)	0.239188	1.046124	0.970452	1.127695
LN Status	**0.000001**	1.908038	1.539252	2.365181
RECQL5 mRNA expression	0.635591	1.078023	0.790092	1.470886

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

Table 3. RECQL5 – RAD51 protein co-expression and breast cancer

VARIABLE	RECQL5 and RAD51 protein co-expression				P- value (Unadjusted)	P -Value (Adjusted)
	RQ5n-/RAD51n- N (%)	RQ5n+/RAD51n- N (%)	RQ5n-/RAD51n+ N (%)	RQ5n+/RAD51n+ N%		
A) Pathological Parameters						
Tumour Size ≤1cm >1-2cm >2-5cm >5cm	12 (6.8) 77 (43.8) 84 (47.7) 3 (1.7)	9 (5.8) 65 (41.9) 78 (50.3) 3 (1.9)	10 (7.6) 73 (55.7) 45 (34.4) 3 (2.3)	16 (8.6) 107 (57.5) 62 (33.3) 1 (0.5)	**0.031**	**0.045**
Tumour Stage 1 2 3	96 (54.5) 64 (36.4) 16 (9.1)	83 (53.2) 57 (36.5) 16 (10.3)	79 (60.3) 44 (33.6) 8 (6.1)	112 (60.2) 52 (28.0) 22 (11.8)	0.370	0.401
Tumour Grade G1 G2 G3	13 (7.4) 59 (33.5) 104 (59.1)	18 (11.5) 38 (24.4) 100 (64.1)	14 (10.7) 53 (40.5) 64 (48.9)	37 (19.9) 73 (39.2) 76 (40.9)	**4.1×10^{-5}**	**0.0005**
Mitotic Index M1 (low; mitoses < 10) M2 (medium; mitoses 10-18) M3 (high; mitosis >18)	45 (26.0) 38 (22.0) 90 (52.0)	34 (22.5) 28 (18.5) 89 (58.9)	41 (31.5) 24 (18.5) 65 (50.0)	72 (40.0) 38 (21.1) 70 (38.9)	**0.008**	**0.017**
Tubule Formation 1 (>75% definite tubule) 2 (10%-75% definite tubule) 3 (<10% definite tubule)	5 (2.9) 45 (26.0) 123 (71.1)	2 (1.3) 49 (32.5) 100 (66.2)	7 (5.4) 43 (33.1) 80 (61.5)	7 (3.9) 75 (41.7) 98 (54.4)	**0.025**	**0.046**

Pleomorphism					**0.027**	**0.043**
1 (small-regular uniform)	0 (0.0)	1 (0.7)	2 (1.6)	2 (1.1)		
2 (Moderate variation)	61 (35.5)	42 (27.8)	48 (37.2)	82 (45.6)		
3 (Marked variation)	111 (64.5)	108 (71.5)	79 (61.2)	96 (53.3)		
Tumour Type					**0.002**	**0.005**
IDC-NST	121 (69.9)	113 (72.4)	75 (58.1)	102 (55.1)		
Tubular Carcinoma	21 (12.1)	22 (14.1)	27 (20.9)	50 (27.0)		
Medullary Carcinoma	7 (4.0)	2 (1.3)	3 (2.3)	0 (0.0)		
ILC	14 (8.1)	8 (5.1)	13 (10.1)	20 (10.8)		
Others	3 (1.7)	1 (0.6)	0 (0.0)	1 (0.5)		
Lymph Node Status					0.353	0.417
Negative	81 (54.4)	69 (51.1)	72 (60.0)	97 (58.4)		
Positive (1-3)	56 (37.6)	54 (40.0)	44 (36.7)	53 (31.9)		
Positive (>3)	12 (8.1)	12 (8.9)	4 (3.3)	16 (9.6)		
B) Aggressive Phenotype						
Her2 overexpression					0.105	0.136
No	153 (87.4)	123 (80.4)	105 (81.4)	159 (88.3)		
Yes	22 (12.6)	30 (19.6)	24 (18.6)	21 (11.7)		
Triple Negative Phenotype					0.457	5.94
No	149 (84.7)	130 (83.3)	115 (87.8)	151 (81.2)		
Yes	27 (15.3)	26 (16.7)	16 (12.2)	35 (18.8)		
NPI					**1.1x10^{-4}**	**0.0005**
≤3.4	31 (18.3)	31 (21.2)	34 (27.4)	68 (38.4)		
>3.4	138 (81.7)	115 (78.8)	90 (72.6)	109 (61.6)		
C) Hormone Receptors						
ER					**7.6x10^{-5}**	**0.0005**
Negative	60 (35.1)	53 (34.9)	32 (25.4)	28 (15.7)		
Positive	111 (64.9)	99 (65.1)	94 (74.6)	150 (84.3)		
PgR					**0.001**	**0.003**
Negative	91 (54.5)	77 (52.0)	49 (39.5)	63 (35.6)		
Positive	76 (45.5)	71 (48.0)	75 (60.5)	114 (64.4)		

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.

Table 4. Multivariate analysis of RECQL5-RAD51 protein co-expression in breast cancer

	B	P value	Exp (B)	95.0% CI for Exp (B)	
				Lower	Upper
Breast Cancer Specific Survival in the whole cohort					
RECQL5-RAD51 co-expression	-0.069	**0.022**	0.934	0.880	0.990
NPI	0.769	**0.001**	2.158	1.378	3.379
ER Status	0.293	0.097	1.340	.949	1.893
HER2 Status	0.728	**0.0001**	2.071	1.432	2.994
Breast Cancer Specific Survival in the ER+ cohort					
RECQL5-RAD51 co-expression	-0.063	0.073	0.939	0.876	1.006
NPI	0.784	**0.001**	2.190	1.376	3.484
HER2 Status	0.848	**0.001**	2.335	1.448	3.767
Breast Cancer Specific Survival in the ER- cohort					
RECQL5-RAD51 co-expression	-0.081	0.172	0.923	0.822	1.036
NPI	0.651	0.519	1.917	0.265	13.882
HER2 Status	0.551	0.057	1.735	0.984	3.062

Figure legends

Figure 1. **A.** *RECQL5* mRNA expression in breast cancer cell lines. **B.** Kaplan Meier curves showing BCSS (Breast cancer specific survival) based on *RECQL5* mRNA expression in the whole cohort **C.** Kaplan Meier curves showing BCSS (Breast cancer specific survival) based on *RECQL5* mRNA expression in ER+ cohort; **D.** Kaplan Meier curves showing BCSS (Breast cancer specific survival) based on *RECQL5* mRNA expression in ER+ cohort that had endocrine therapy; **E.** Kaplan Meier curves showing BCSS (Breast cancer specific survival) based on *RECQL5* mRNA expression in ER- cohort. **F.** Kaplan Meier curves showing BCSS (Breast cancer specific survival) based on *RECQL5* mRNA expression in ER- cohort that had chemotherapy.

Figure 2. A1. Western blot of RECQL5 expression in breast cancer cell lines. **A2**. MCF10A breast cells were transfected with RECQL5 or empty vector and. Extracts were made 48 h post transfection and western blotted for RECQL5. RECQL5 level in MCF7 cells is shown for comparison. **B.** MCF10A breast cells were transfected with RECQL5 or empty vector and cell number counted every 24 h. Mean and standard deviation are shown for 3 independent repeats. ** indicates a statistical significance of P<0.01 at 72 h using Student's T.Test. **C.** Western blot of RECQL5 expression in MCF-7 (RECQL5 wild-type) and MCF-7 RECQL5 knock down cell line. **D.** Microphotograph of RECQL5 protein expression in breast tumours. **E.** Kaplan Meier curves showing BCSS based on RECQL5-RAD51 co-expression in the whole cohort. **F.** Kaplan Meier curves showing BCSS based on RECQL5-RAD51 co-expression in ER+ cohort.

A

B RECQL5 mRNA Expression in Whole Cohort

Low RECQL5 (n=1306)

High RECQL5 (n=671)

Log Rank = 13.251; **P = 2.7x10^-4**
Monte-Carlo P <0.0001

Breast Cancer Specific Survival(months)

C RECQL5 mRNA Expression in ER+ Cohort

Low RECQL5 (n=976)

High RECQL5 (n=521)

Log Rank = 14.78; **P = 1.2x10^-4**
Monte-Carlo P = 0.01

Breast Cancer Specific Survival(months)

D ER+; NPI>3.4; had endocrine therapy

Low RECQL5 (n=605)

High RECQL5 (n=498)

Log Rank = 11.62 ; **P = 0.001,**
Monte-Carlo P= 0.020

Breast Cancer Specific Survival(months)

E RECQL5 mRNA Expression in ER- cohort

Low RECQL5 (n=204)

High RECQL5 (n=234)

Log Rank =2.473 ; P = 0.116,
Monte-Carlo P= 0.7700

Breast Cancer Specific Survival(months)

F ER-; NPI>3.4; had chemotherapy

Low RECQL5 (n=120)

High RECQL5 (n=152)

Log Rank = 1.55; P = 0.213
Monte-Carlo P= 0.9300

Breast Cancer Specific Survival(months)

Figure 1

A

MCF-7
MDA-MB231
MDA-MB-436
MDA-MB-468

RECQL5

β-Actin

B

MCF-7
MCF10A
MCF10A + Q5

RECQL5

γ-Tubulin

C

◆ MCF10A
◻ MCF10A + Q5

**

6.E+04
5.E+04
3.E+04
2.E+04
1.E+00

Cell number

0   18   36   54   72

Hours post transfection

D

MCF-7 (control)
MCF-7 (RECQL5 KD 1)
MCF-7 (RECQL5 KD 2)

RECQL5

β-Actin

E

RECQL5 negative

RECQL5 positive

F

RECQL5-RAD51 co-expression
and survival (Whole cohort)

1.0
0.8
0.6
0.4
0.2
0.0

Cumulative Survival

RECLQ5+/RAD51+ (n=142)
RECQL5-/RAD51+ (n=102)
RECQL5-/RAD51n (n=126)
RECQL5+/RAD51- (n=115)

Log Rank = 13.796; P=0.003

0   50   100   150   200

Breast Cancer Specific Survival
(months)

G

RECQL5- RAD51 co-expression
and survival (ER+ Cohort)

1.0
0.8
0.6
0.4
0.2
0.0

Cumulative Survival

RECQL5+/RAD51+ (n=110)
RECQL5-/RAD51+ (n=74)
RECQL5-/RAD51- (n=75)
RECQL5+/RAD51- (n=67)

Log Rank = 11.395; P = 0.010

0   50   100   150   200

Breast Cancer Specific Survival
(months)

Figure 2

Supplementary Figure Legends

Supplementary Figure S1: A. Kaplan Meier curves showing BCSS based on RECQL5-expression in the whole cohort. **B.** Kaplan Meier curves showing BCSS based on RECQL5-expression in the ER+ cohort. **A.** Kaplan Meier c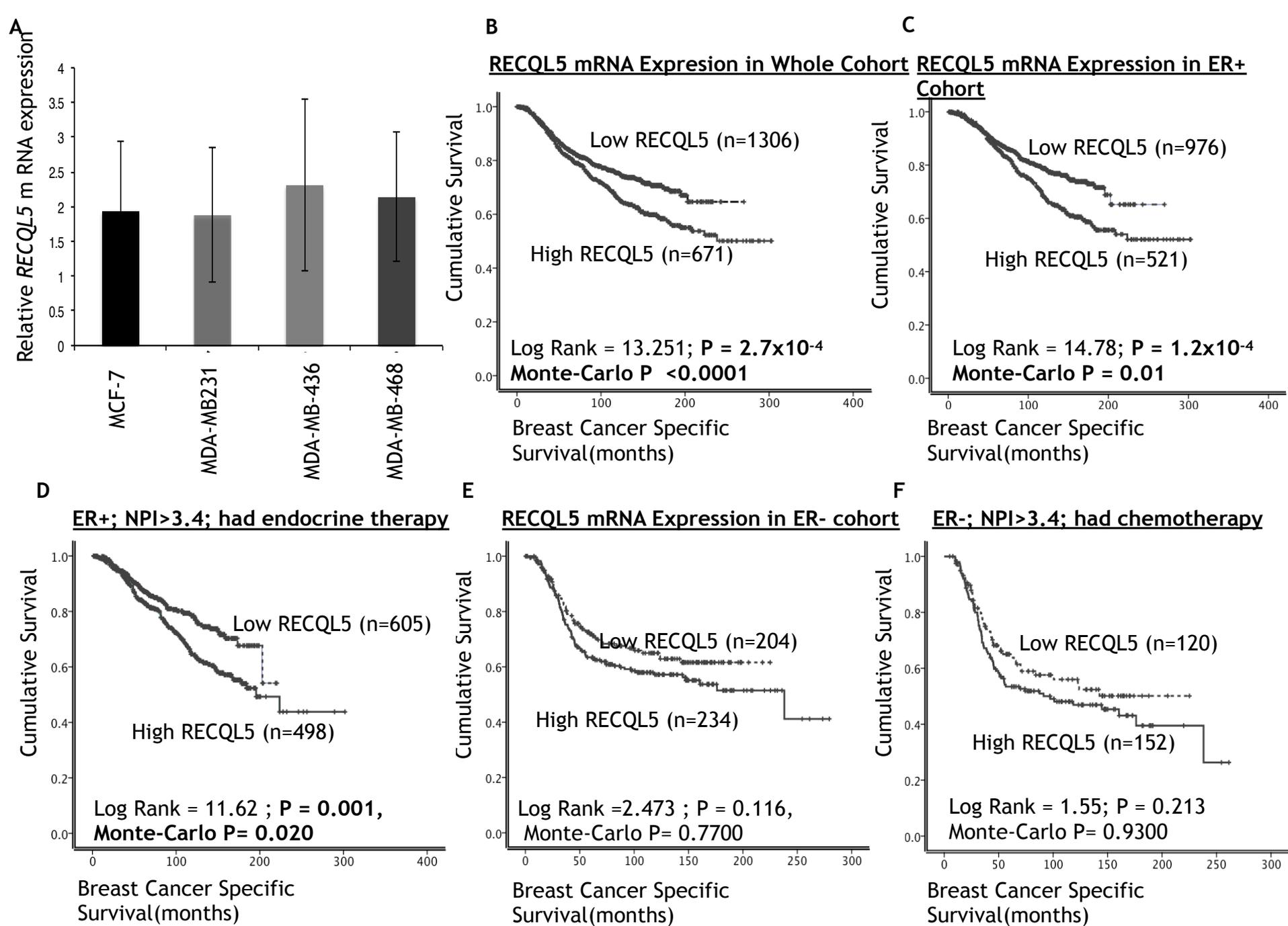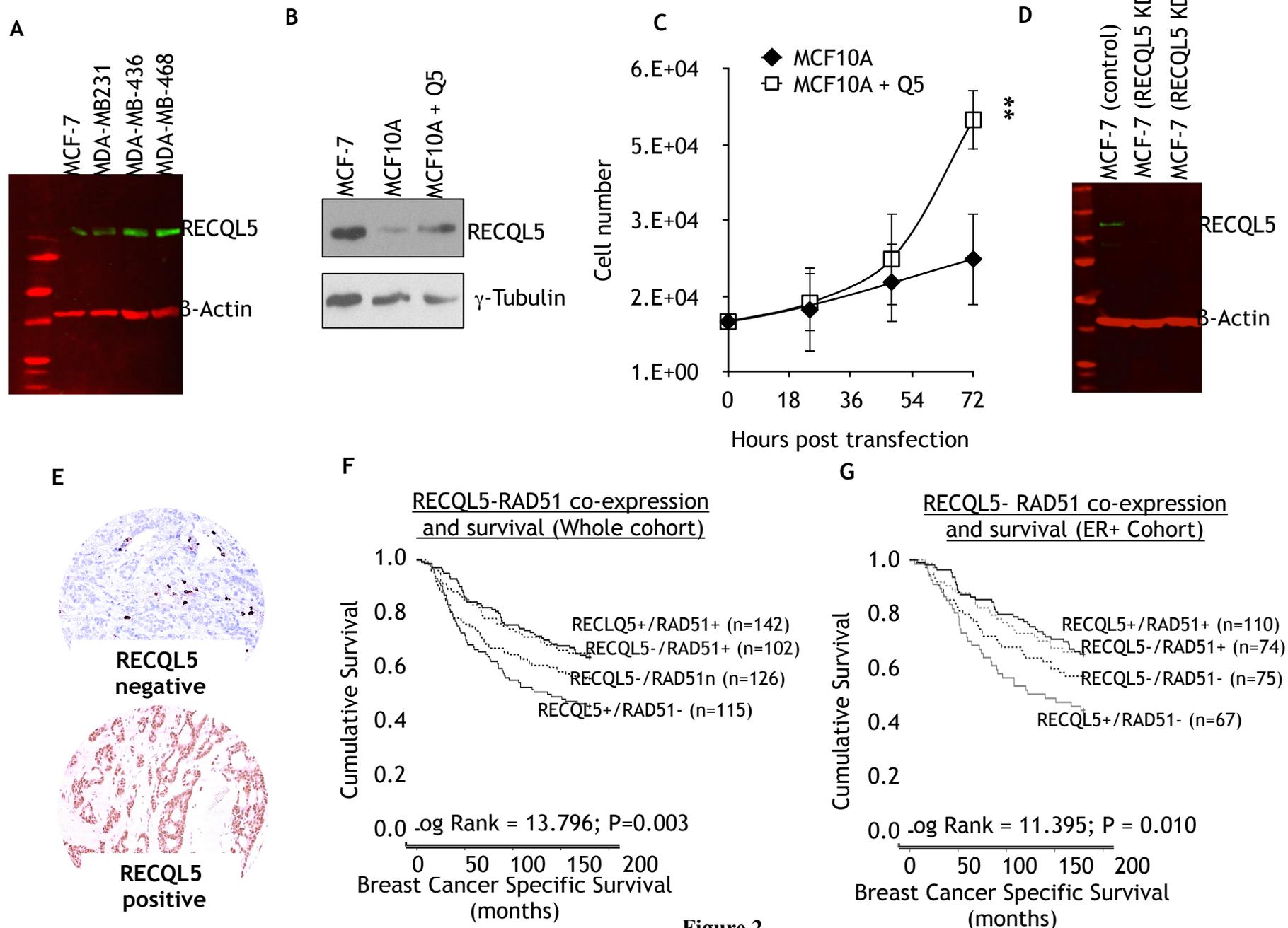urves showing BCSS based on RECQL5-expression in the ER- cohort. **D.** Kaplan Meier curves sho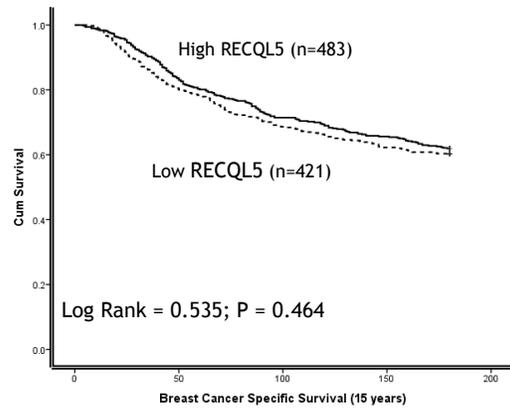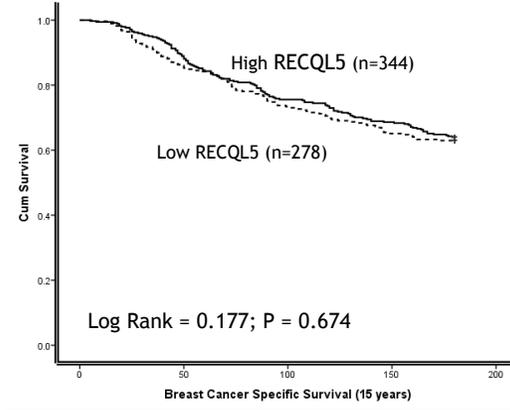wing BCSS based on RECQL5-RAD51 co-expression in ER- cohort. **D.** Kaplan Meier curves showing BCSS based on RECQL5-TOPO2A co-expre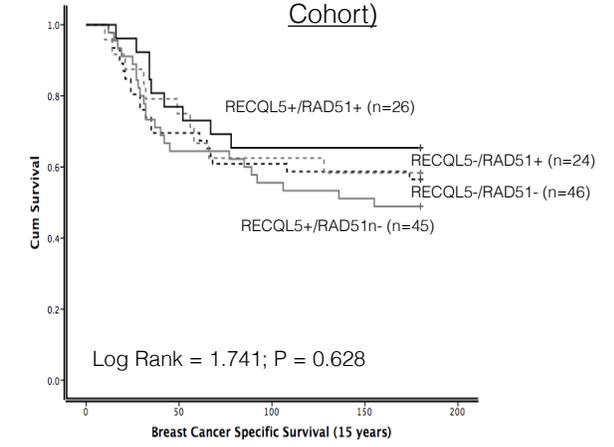ssion in whole cohort. **D.** Kaplan Meier curves showing BCSS based on RECQL5-TOPO2A co-expression in ER+ cohort. **D.** Kaplan Meier curves showing BCSS based on RECQL5-TOPO2A co-expression in ER- cohort.
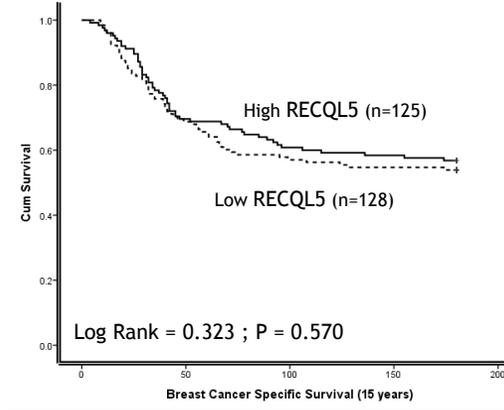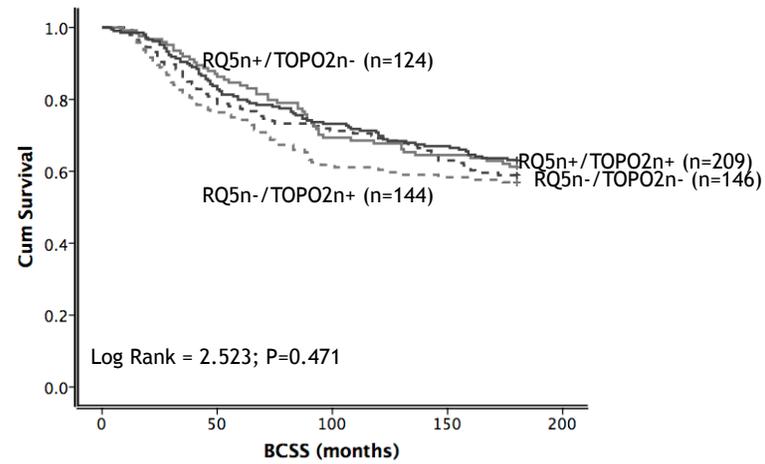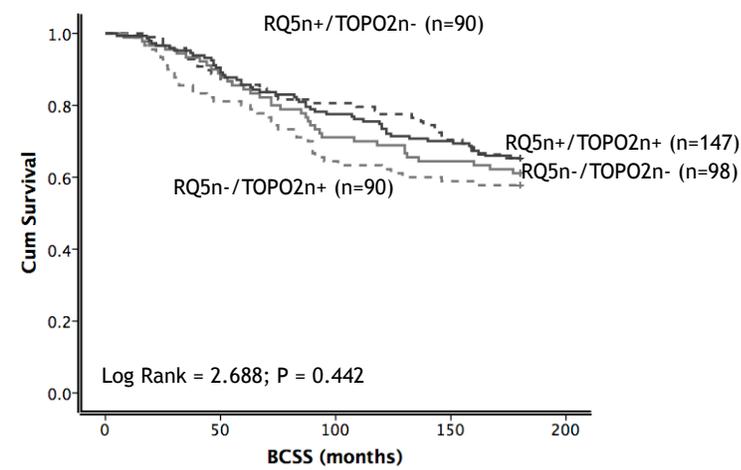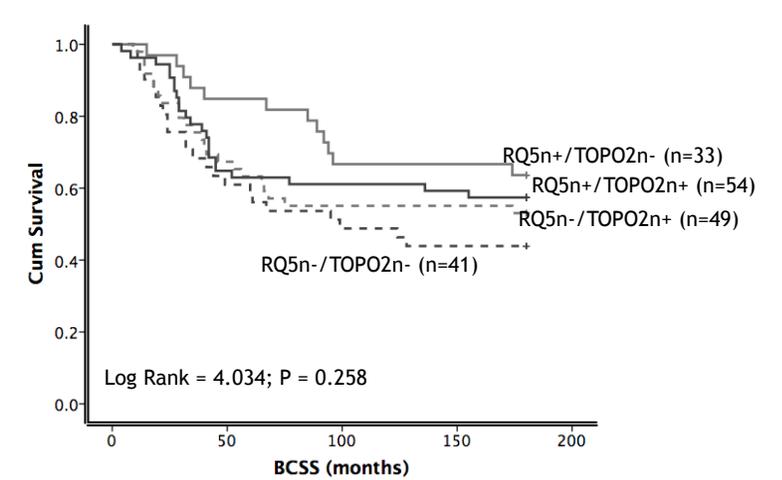
Supplementary Figure S1

Supplementary Table S1: Clinicopathological characteristics in the METABRIC external validation 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 1 2 3 >3	 1012 336 170 112 316
ER status Positive Negative	 1485 437
PAM50 subtype Basal HER2 Luminal A Luminal B Normal Not classified	 322 238 714 484 188 6
Adjuvant systemic therapy (AT)	
No AT	290
Hormone therapy (HT)	1014
Chemotherapy	226
Hormone + chemotherapy	192

Supplementary Table S2: Clinicopathological characteristics of Nottingham Tenovus series

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

Supplemental Table S3: Clinicopathological characteristics of ER- cohort

Variable	n*	Cases	(%)
Menopausal status	252		
Pre-menopausal		122	(48.5)
postmenopausal		130	(51.5)
Tumour Grade (NGS)	252		
G1		1	(0.3)
G2		27	(10.6)
G3		224	(89.1)
Lymph node stage	252		
Negative		121	(48)
Positive (1-3 nodes)		86	(34)
Positive (>3 nodes)		45	(18)
Tumour size (cm)	252		
T1 a + b (≤1.0)		28	(11)
T1 c (>1.0 -2.0)		106	(42)
T2 (>2.0-5)		103	(41)
T3 (>5)		15	(6)
Tumour type	252		
IDC-NST		224	(89.0)
Tubular		5	(2.0)
ILC		8	(3.0)
Medullary (typical/atypical)		5	(2.0)
Others		0	(4.0)
NPI subgroups	252		
Excellent PG(2.08-2.40)	Low risk	0	(0.0)
Good PG(2.42-3.40)		0	(0.0)
Moderate I PG(3.42 to 4.4)	High risk	111	(44.0)
Moderate II PG(4.42 to 5.4)		81	(32.0)
Poor PG(5.42 to 6.4)		38	(15.0)
Very poor PG(6.5-6.8)		22	(9.0)
Survival at 5 years	252		
Alive and well		176	(70.0)
Dead from disease		73	(29.0)
Dead from other causes		3	(1.0)

* 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)
ATM	Rabbit MAb anti-ATM	Y170	Abcam	Citrate pH6	1:100 18 hours	Nuclear	% of positive cells	<25% (negative)
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)
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	Nuclear	H-score	< 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)
Ck6	Mouse MAb anti-Ck5/6	D5/161 B4	Dako-Cytomation	EDTA pH8	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	Nuclear/ cytoplasm	% of positive cells	>25% (positive)
RECQL5	Rabbit anti RecqL5	Polyclonal	Sigma-Aldrich	Citrate pH6	1:100 60 min	Nuclear	H - Score	≥10 (positive)
BLM	Rabbit anti BLM	Polyclonal	Novus-Biologicals	Citrate pH6	1:100 18 Hours	Nuclear/ Cytoplasmic	H- Score	≥50 (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: RecQL5 and breast cancer

VARIABLE	RecQL5 (Nuclear) Protein Expression		P- values (Unadjusted)	P –Values (Adjusted)
	Low N (%)	High N (%)		
A) Pathological Parameters				
Tumour Size <1cm >1-2cm >2-5cm >5cm	38 (6.9) 271 (49.0) 230 (41.6) 14 (2.5)	69 (10.7) 325 (50.6) 235 (36.6) 13 (2.0)	0.061	0.09
Tumour Stage 1 2 3	326 (58.8) 183 (33.0) 45 (8.1)	400 (62.2) 181 (28.1) 62 (9.6)	0.161	0.1982
Tumour Grade G1 G2 G3	67 (12.1) 182 (32.9) 304 (55.0)	122 (19.0) 209 (32.6) 311 (48.4)	**0.003**	**0.008**
Mitotic Index M1 (low; mitoses < 10) M2 (medium; mitoses 10-18) M3 (high; mitosis >18)	172 (31.3) 97 (17.6) 281 (51.5)	233 (36.5) 124 (19.4) 281 (44.0)	**0.050**	0.075
Tubule Formation 1 (>75% of definite tubule) 2 (10%-75% definite tubule) 3 (<10% definite tubule)	27 (4.9) 155 (28.2) 368 (66.9)	35 (5.5) 239 (37.5) 364 (57.1)	**0.002**	**0.005**
Pleomorphism 1 (small-regular uniform) 2 (Moderate variation) 3 (Marked variation)	12 (2.2) 194 (35.3) 344 (62.5)	18 (2.8) 245 (38.4) 375 (58.8)	0.377	0.431
Tumour Type IDC-NST Tubular Carcinoma Medullary Carcinoma ILC Others	309 (63.4) 73 (15.0) 16 (3.3) 38 (7.8) 51 (10.5)	322 (59.9) 51 (22.1) 6 (1.1) 13 (9.1) 26 (7.8)	**0.004**	**0.009**
Lymphovascular Invasion No Yes	361 (65.9) 187 (34.1)	421 (66.2) 215 (33.8)	0.908	43.58
B) Aggressive phenotype				
Her2 overexpression No Yes	466 (86.3) 74 (13.7)	567 (90.6) 59 (9.4)	**0.022**	**0.036**
Triple Negative Phenotype No Yes	420 (78.1) 118 (21.9)	516 (82.8) 107 (17.2)	**0.041**	0.065

Basal Like Phenotype No Yes	446 (85.3) 77 (14.7)	532 (89.3) 64 (10.7)	**0.045**	0.069
Cytokeratin 6 (CK6) Negative Positive	391 (82.1) 85 (17.9)	440 (84.1) 83 (15.9)	0.402	0.448
Cytokeratin 14 (CK14) Negative Positive	410 (87.2) 60 (12.8)	451 (86.7) 69 (13.3)	0.814	0.831
Cytokeratin 18 (CK18) Negative Positive	54 (12.2) 388 (87.8)	44 (9.1) 442 (90.9)	0.117	0.161
Cytokeratin 19 (CK19) Negative Positive	36 (7.6) 440 (92.4)	27 (5.2) 495 (94.8)	0.121	0.161
C) Hormone receptors				
ER Negative Positive	171 (31.3) 375 (68.7)	150 (23.8) 480 (76.2)	**0.004**	**0.008**
PgR Negative Positive	245 (46.9) 483 (53.1)	249 (42.0) 344 (58.0)	0.097	0.137
AR Negative Positive	191 (42.4) 259 (57.6)	166 (33.7) 327 (66.3)	**0.006**	**0.011**
D) DNA Repair				
BRCA1 Absent Normal	103 (26.0) 293 (74.0)	67 (15.1) 377 (84.9)	**8.4×10^{-6}**	**0.00001**
XRCC1 Low High	91 (22.1) 320 (77.9)	49 (10.7) 408 (89.3)	**5.0×10^{-6}**	**0.00001**
FEN1 Low High	292 (78.5) 80 (21.5)	295 (69.1) 132 (30.9)	**0.003**	**0.007**
SMUG1 Low High	176 (48.4) 188 (51.6)	141 (33.6) 279 (66.4)	**2.6×10^{-5}**	**0.0001**
APE1 Low High	292 (63.8) 166 (36.2)	213 (39.7) 324 (60.3)	**1.0×10^{-6}**	**0.00001**

Pol β Low High	221 (45.9) 261 (54.1)	189 (33.3) 378 (66.7)	**3.5×10^{-5}**	**0.0001**
ATR Low High	344 (74.1) 120 (25.9)	315 (60.7) 204 (39.3)	**8.0×10^{-6}**	**0.00001**
ATM Low High	195 (57.4) 145 (42.6)	187 (48.4) 199 (51.6)	**0.016**	**0.027**
DNA-PKcs Low High	200 (45.9) 236 (54.1)	133 (26.7) 365 (73.3)	**1.0×10^{-6}**	**0.00001**
TOP2A Low Overexpression	204 (53.1) 180 (46.9)	164 (37.6) 272 (62.4)	**8.0×10^{-6}**	**0.00001**
pCHK1 (Cytoplasmic) Low High	214 (38.6) 341 (61.4)	188 (29.2) 456 (70.8)	**0.001**	**0.003**
CHK2 Low High	230 (57.2) 172 (42.8)	168 (38.4) 270 (61.6)	**1.0×10^{-6}**	**0.00001**
RecQL4 (Nuclear) Low High	279 (71.2) 113 (28.8)	209 (47.3) 233 (52.7)	**2.7×10^{-12}**	**0.00001**
RecQL4 (Cytoplasmic) Low High	230 (58.8) 161 (41.2)	204 (46.4) 236 (53.6)	**0.00033**	**0.001**
BLM (Nuclear) Low High	170 (37.0) 290 (63.0)	85 (16.5) 429 (83.5)	**4.5×10^{-13}**	**0.00001**
BLM (Cytoplasmic) Low High	355 (77.2) 105 (22.8)	355 (16.5) 155 (30.4)	**0.008**	**0.014**
RAD51 (Nuclear) Low High	176 (57.3) 131 (42.7)	156 (45.6) 186 (54.4)	**0.003**	**0.007**
RAD51 (Cytoplasmic) Low High	15 (4.9) 292 (95.1)	25 (7.4) 311 (92.6)	0.180	0.211

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 les

Supplementary Table S6: RECQL5 – TOPO2 co-expression and breast cancer

VARIABLE	RecQL5 (Nuclear) & TOPO2(Nuclear) Protein Co-Expression				P- value (Unadjusted)	P- value (Adjusted)
	RQ5n-/ TOPO2n- N (%)	RQ5n+/ TOPO2n- N (%)	RQ5n-/ TOPO2n+ N (%)	RQ5n+/ TOPO2n+ N%		
Tumour Size ≤1cm >1-2cm >2-5cm >5cm	15 (7.2) 103 (49.3) 88 (42.1) 3 (1.4)	13 (7.9) 76 (46.3) 72 (43.9) 3 (1.8)	10 (5.5) 90 (49.5) 78 (42.9) 4 (2.2)	25 (9.1) 144 (52.6) 103 (37.6) 2 (0.7)	0.743	9.659
Tumour Stage 1 2 3	126 (60.3) 72 (34.4) 11 (5.3)	98 (59.4) 50 (30.3) 17 (10.3)	108 (59.3) 51 (28.0) 23 (12.6)	171 (62.2) 78 (28.4) 26 (9.5)	0.231	0.301
Tumour Grade G1 G2 G3	31 (14.8) 79 (37.8) 99 (47.4)	30 (18.3) 51 (31.1) 83 (50.6)	14 (7.7) 44 (24.2) 124 (68.1)	51 (18.5) 90 (32.7) 134 (48.7)	**2.4X10^{-4}**	**.001**
Mitotic Index M1 (low; mitoses < 10) M2 (medium; mitoses 10-18) M3 (high; mitosis >18)	66 (32.2) 44 (21.5) 95 (46.3)	60 (37.5) 21 (13.1) 79 (49.4)	43 (24.0) 26 (14.5) 110 (61.5)	101 (37.4) 47 (17.4) 122 (45.2)	**0.006**	**0.015**
Tubule Formation 1 (>75% definite tubule) 2 (10%-75% definite tubule) 3 (<10% definite tubule)	10 (4.9) 78 (38.0) 117 (57.1)	9 (5.6) 66 (41.3) 85 (53.1)	4 (2.2) 35 (19.6) 140 (78.2)	11 (4.1) 97 (35.9) 162 (60.0)	**9.1X10^{-5}**	**0.0012**
Pleomorphism 1 (small-regular uniform) 2 (Moderate variation) 3 (Marked variation)	4 (2.0) 82 (40.0) 119 (58.0)	1 (0.6) 56 (35.0) 103 (64.4)	1 (0.6) 50 (28.2) 126 (71.2)	5 (1.9) 104 (38.5) 161 (59.6)	0.121	0.174
Tumour Type IDC-NST Tubular Carcinoma Medullary Carcinoma ILC Others Mixed NST/Lobular/Special Type	126 (61.2) 45 (21.8) 5 (2.4) 12 (5.8) 3 (1.5) 15 (7.3)	105 (64.4) 36 (22.1) 3 (1.8) 10 (6.1) 2 (1.2) 7 (4.3)	128 (71.9) 10 (5.6) 9 (5.1) 21 (11.8) 2 (1.1) 8 (4.5)	166 (61.0) 67 (24.6) 2 (0.7) 25 (9.2) 3 (1.1) 9 (3.3)	**1.4X10^{-4}**	**0.0009**

Lymph Node Status Negative Positive (1-3) Positive (>3)	107 (57.8) 69 (37.3) 9 (4.9)	88 (57.9) 51 (33.6) 13 (8.6)	99 (60.7) 47 (28.8) 17 (10.4)	149 (61.1) 73 (29.9) 22 (9.0)	0.364	0.430
Her2 overexpression No Yes	175 (85.8) 29 (14.2)	136 (85.5) 23 (14.5)	145 (80.1) 36 (19.9)	243 (89.0) 30 (11.0)	0.072	0.134
Triple Negative Phenotype No Yes	171 (81.8) 38 (18.2)	144 (87.3) 21 (12.7)	150 (82.4) 32 (17.6)	223 (81.1) 73 (18.9)	0.385	0.417
NPI ≤3.4 >3.4	58 (28.6) 145 (71.4)	47 (30.9) 105 (69.1)	31 (18.5) 137 (81.5)	96 (36.5) 167 (63.5)	**0.001**	**0.003**
ER Negative Positive	58 (28.7) 144 (71.3)	42 (25.9) 120 (74.1)	60 (34.1) 116 (65.9)	63 (23.7) 203 (76.3)	0.107	0.174
PgR Negative Positive	86 (42.4) 114 (57.6)	72 (47.1) 81 (52.9)	87 (51.2) 83 (48.8)	100 (37.7) 165 (62.3)	**0.036**	**0.038**

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.