Evaluation of CDK12 Protein Expression as a Potential Novel Biomarker for DNA Damage Response Targeted Therapies in Breast Cancer

Kalnisha Naidoo[1], Patty T. Wai[1,2], Sarah L. Maguire[1], Frances Daley[1], Syed Haider[1] Divya Kriplani[1], James Campbell[1,3], Hasan Mirza[4], Anita Grigoriadis[4], Andrew Tutt[1,5], Paul M. Moseley[6], Tarek M.A. Abdel-Fatah[6], Stephen Y. Chan[6], Srinivasan Madhusudan[7], Emad A. Rhaka[7], Ian O. Ellis[7], Christopher J. Lord[1,3], Yinyin Yuan[2], Andrew R. Green[7] and Rachael Natrajan[1,2*]

1. The Breast Cancer Now Research Centre, The Institute of Cancer Research, London, UK.

2. Division of Molecular Pathology, Centre for Evolution and Cancer and Centre for Molecular Pathology, The Institute of Cancer Research, London

3. The CRUK Gene Function Laboratory, The Institute of Cancer Research, London, UK.

4. Cancer Bioinformatics, Cancer Division, King's College London, London, UK

5. Breast Cancer Now Research Unit, King's College London, London, UK

6. Clinical Oncology, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, UK.

7. Department of Histopathology and Division of Cancer & Stem Cells, School of Medicine, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, UK.

Running title: CDK12 expression as a biomarker in breast cancer

Keywords: CDK12; breast cancer; DNA repair; PARP inhibitors

Financial support: This study was funded by Breast Cancer Now. RN is the recipient of a Breast Cancer Now Career Development Fellowship (2011MaySF01).

Corresponding author:

*Rachael C. Natrajan, PhD, The Breast Cancer Now Toby Robins Research Centre, Mary-Jean Mitchell Green Building, The Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP. Email: rachael.natrajan@icr.ac.uk

Word count: (4091)

Figures: 4, Tables: 1

Abstract:

Disruption of Cyclin Dependent Kinase 12 (*CDK12*) is known to lead to defects in DNA repair and sensitivity to platinum salts and poly(ADP-ribose) polymerase 1/2 inhibitors. However, *CDK12* has also been proposed as an oncogene in breast cancer. We therefore aimed to assess the frequency and distribution of CDK12 protein expression by immunohistochemistry (IHC) in independent cohorts of breast cancer and correlate this with outcome and genomic status. We found that 21% of primary unselected breast cancers were CDK12 high, and 10.5% were absent, by IHC. CDK12 positivity correlated with HER2 positivity but was not an independent predictor of breast cancer specific survival taking HER2 status into account, however absent CDK12 protein expression significantly correlated with a triple negative phenotype. Interestingly, CDK12 protein absence was associated with reduced expression of a number of DDR proteins including ATR, Ku70/Ku80, PARP1, DNA-PK and gamma-H2AX, suggesting a novel mechanism of CDK12 associated DDR dysregulation in breast cancer. Our data suggest that diagnostic IHC quantification of CDK12 in BC is feasible, with CDK12 absence possibly signifying defective DDR function. This may have important therapeutic implications, particularly for triple negative breast cancers.

Introduction

Breast cancer (BC) is a complex disease comprising a variety of molecular and clinically distinct subtypes. Substantial progress has been made in the management of BC mortality over the last 25 years, in part due to improved treatment modalities such as endocrine therapies, HER2-targeted therapy and combination chemotherapies (1-4). However, a proportion of sporadic primary BC remain difficult to treat. Hence, there is an urgent need for stratification and biomarker discovery within this cohort.

The CycK/CDK12 (Cyclin K/Cyclin dependent kinase 12) complex is involved in the regulation of RNA polymerase II (RNA pol II) and mRNA processing (5-7) and is known to protect normal cells from genomic instability by regulating the transcription of DNA damage response (DDR) genes (8). Moreover, CDK12 has been postulated as a tumor suppressor gene in high-grade serous ovarian cancer (HGSOC), where it is one of the only significantly recurrently mutated genes (9). Recurrent point mutations have been shown to abrogate the functional activity of CDK12, resulting in defects in multiple DNA repair pathways, leading to genomic instability, down-regulation of some homologous recombination (HR) genes such as BRCA1, FANCI or FANCD2 (10,11) and selective sensitivity to both platinum agents and poly (ADP-ribose) polymerase (PARP1/2) inhibitors (12,13). Indeed, recent data in HGSOC suggests that CDK12 inactivated tumors have a unique signature of genomic instability characterized by frequent mega-sized gains scattered over the genome, that is a result of numerous tandem duplications, indicative of gross defects in DNA repair (14). In addition, recent profiling studies have also identified *CDK12* mutations in primary and castration resistant prostate cancer that are mutually exclusive with other mutations in DNA repair genes (15,16), and akin to HGSOC, result in large tandem duplications (14). On the other hand, in BC, *CDK12* gene amplification often

co-occurs with *ERBB2* amplification as both are co-located at locus Ch17q12 (17,18), and CDK12 overexpression has been correlated with indicators of aggressive disease, suggesting that CDK12 could act as a oncogenic driver and prognostic biomarker in BC as a result of this co-location (19).

We have previously shown that in BC, *CDK12* is recurrently targeted by both DNA rearrangements (13% of HER2-amplified BC) and recurrent point mutations (2.6% of unselected BC) (13) in a similar manner to HGSOC, and that loss of CDK12 in BC models confers sensitivity to PARP1/2 inhibitors *in vitro* through defects in HR (12,13). Loss of CDK12 in BC may therefore signify response to platinum salts and/or PARP1/2 inhibitors (12,13).

Here we sought to i) investigate the distribution and frequency of CDK12 protein expression in a large series of unselected and Herceptin treated HER2-positive BC, using immunohistochemistry (IHC) and examine any correlation with survival; ii) evaluate CDK12 protein and mRNA expression with genomic alterations and iii) assess whether CDK12 would constitute an oncogenic driver in *CDK12* amplified tumors.

Materials and Methods

Tissue Microarray Patient Cohorts

Unselected BC

Primary operable BC cases (n= 1,650) from the Nottingham Tenovus Primary Breast Carcinoma Series were utilized as previously described (20-22). Patients were under the age of 71 years (median, 55 years), diagnosed between 1986 and 1999, and

treated uniformly in a single institution. Clinicopathological parameters for this series are summarized in Supplementary Table S1.

HER2-positive adjuvant trastuzumab series

The HER2-positive adjuvant trastuzumab series comprises 143 primary operable BC from patients presenting between 2003 and 2010 who received adjuvant trastuzumab (21). HER2 status was determined according to the American Society of Clinical Oncology (ASCO) guidelines as previously described (21). Clinicopathological parameters for this series are summarized in Supplementary Table S2.

METABRIC Nottingham Breast Cancers

This series comprised 282 primary BC from Nottingham, which form part of the METABRIC cohort (23), (Supplementary Table S3).

Tissue Microarray (TMA) Construction

Tumor samples were arrayed as previously described (22). Briefly, one core per tumor of 0.6 mm thickness was obtained from the most representative areas then re-embedded in microarray blocks.

CDK12 Immunohistochemistry

IHC was optimized in-house, using a standard Labelled Polymer technique, on 4µm sections of formalin-fixed paraffin embedded (FFPE) normal human tonsil; cell blocks containing the MCF7 breast cancer cell line known to express CDK12 transfected with a previously validated siRNA pool targeting CDK12 or non-targeting control and BT474 cells as a positive control (13) (Fig. 1). Cells were cultured as previously described (13) and authenticated by short tandem repeat (STR) typing using the StemElite Kit (Promega, UK). Briefly, slides were dewaxed in xylene and rehydrated

through graded alcohols. Following heat-induced antigen retrieval in citrate buffer (pH 6.0), sections were incubated with a mouse anti-human CDK12 monoclonal antibody (1:5000 final dilution, Abcam clone 57311 that was raised against an immunogen peptide corresponding to amino acids 1281-1380 of Human CDK12) for one hour at room temperature. The staining was visualized using the Dako Flex Envision K8002 Kit (Dako), counterstained with Gills hematoxylin (Leica). Sections were then dehydrated and mounted.

TMAs were assessed for nuclear CDK12 protein expression in the malignant epithelium only, using a modified Allred score (14). Only technically sound cores containing >20% invasive tumor cells were included in the analysis. Cores were evaluated for both intensity (0 = no stain; 1 = mild; 2 = moderate; 3 = strong), and percentage of epithelial cells that stained positive (0 = absent; 1 = background; 2 = 1–25%; 3 = 26–50%; 4 = 51-75%; 5 = > 75%), Fig. 1. Scores were derived from a sum of the intensity and percentage of immunoreactive cells; an average score of 0 for each tumor was considered negative/absent, and a score of 7 or 8, high, and a score of 2-6 as intermediate expression. Scores of 1 were excluded from further analysis as these equated to background non-specific staining. IHC staining and dichotomization of the other biomarkers included in this study were as per previous publications (24-30). Scoring was performed blinded to the study endpoint.

Mining of public datasets

In order to corroborate our findings, we re-analysed publicly available data from The Cancer Genome Atlas (31,32) and METABRIC (23) datasets, to ascertain the frequency of *CDK12* copy number breakpoints, somatic mutations and methylation and correlate these with RNA expression levels. Low and high CDK12 gene expression were defined by using an optimal threshold for dichotomizing gene expression data as described (33). This was carried out by a stepwise analysis from

40 to 60 percentiles at an interval of 5. The cut-offs that displayed the highest prognostic significance with log-rank test were selected. In addition, analysis of published whole genome shRNA (34) and kinome wide siRNA (35) genetic perturbation screens was performed to correlate cell viability of breast cancer cell lines with and without *CDK12* amplification after CDK12 knockdown.

Assessment of Tandem Duplicator Phenotype

Affymetrix SNP6.0 copy number data of 224 METABRIC samples were preprocessed using PennCNV-affy package (affy: http://penncnv.openbioinformatics.org/en/latest/misc/credit/ and segmented absolute copy number and ploidy was established with ASCAT 2.1 (36). The two tandem duplication phenotypes were established as previously described in Watkins et al., 2016 (37).

Statistical Analysis

Retrospective statistical analysis was performed using SPSS 21.0 statistical software (SPSS Inc., Chicago, IL, USA), in compliance with reporting recommendations for tumor marker prognostic studies (REMARK) criteria (38). A chi-squared or Fisher's exact test of <0.05 was considered significant. Survival curves were analyzed by the method of Kaplan-Meier, with a p-value <0.05 being considered significant with a 10-year BC specific survival as the endpoint. Multivariate survival analysis was carried out using CDK12 expression status, node status, estrogen and progesterone receptor (ER/PR) status, HER2 status, age, tumor size and grade. A Student's *t*-test was employed to compare CDK12 expression of mined samples with genetic aberrations and normal controls. For comparisons, scores of 0 (absent) and high (7-8) were compared, given known correlations with high expression and amplification (19) and uncertainty regarding intermediate levels of expression. A p-value <0.05 (two-sided) was considered statistically significant. Differential gene expression

analysis of CDK12 absent versus high tumors with gene expression known to be involved in DNA repair was performed using data from METABRIC using Limma with FDR (False Discovery Rate) multiple correction adjustment.

Results

Distribution of CDK12 expression and clinicopathological correlation

In the first instance, we assessed CDK12 protein expression by IHC in 696 unselected BC samples that met the inclusion criteria (described in methods; Fig. 1). Overall 73/696 tumors were absent/negative for CDK12 (10.5%) by IHC, and 146 had CDK12 high expression (21%; Fig. 1A, Supplementary Table S1). Breaking this down by subtype, in ER+ patients, 50/510 (9.8%) and 101/510 (19.8%) were CDK12 absent and high; in HER2+ patients 3/102 (2.9%) and 55/102 (53.9%) were CDK12 absent and high; and in TN patients, 21/123 (17%) and 17/123 (13.8%) were CDK12 absent and high respectively. Expression of CDK12 significantly correlated with HER2 expression; 96% of CDK12 absent tumors were HER2 negative and 95% of HER2 positive tumors had high CDK12 expression ($p < 0.001$, Chi-Square test). Interestingly, no significant correlation of CDK12 expression with ER or PR status was observed, but a greater proportion of CDK12 absent tumors showed a triple-negative phenotype (21/73, 29%) than CDK12 high tumors (17/143, 11.9%, $p = 0.002$, Chi-Square test**,** Supplementary Table S1). There was no association with CDK12 high expression and breast cancer specific survival in this cohort ($p = 0.354$, HR = 1.295, 95% CI= 0.75-2.24, Log Rank (Mantel-Cox), Fig. 2A, Supplementary Table S4).

These findings were validated in a subset of tumors from the METABRIC cohort of unselected BC, in which CDK12 was highly expressed in 63/250 tumors (25.2%) and absent in 89/250 tumors (35.6%), Fig. 1 Supplementary Table S3**.** Again, a

significant correlation with HER2 status was observed, with 83/89 (93.3%) of CDK12 absent tumors being HER2 negative and 13/36 (36.1%) of HER2 amplified tumors being CDK12 high ($p < 0.0001$, Chi-Square test). CDK12 expression also conferred a significantly poorer BC specific survival ($p < 0.001$, HR = 3.161, 95% CI = 1.632-6.125, Log Rank (Mantel-Cox)) in this cohort in univariate analysis (Fig. 2B). This was also significant in multivariate comparisons when taking account for HER2 positivity ($p= 0.038$, HR = 2.26, 95% CI = 1.045-4.887, Supplementary Table S4). These associations were further corroborated at the mRNA level in a larger cohort of primary tumors (n=1961) from METABRIC, where high CDK12 expression was associated with a worse BC specific survival ($p < 0.001$, HR = 1.28, 95% CI = 1.17 - 1.41, Log Rank (Mantel-Cox), Fig. 2C), however this was not substantiated in multivariate analysis when taking HER2 into account (Supplementary Table S4)**.**

We previously identified a proportion of HER2 amplified tumors harbour out-of-frame *CDK12* fusion genes. As CDK12 is known to map to the smallest region within the *HER2* amplicon (17,18) these fusions are the result of an amplification breakpoint in the *HER2* amplicon that converges on CDK12, disrupting its expression (13). We also found that tumor cells with loss of CDK12 due to the presence of a breakpoint in the *HER2* amplicon were sensitive to PARP1/2 inhibitors suggesting that a fraction of HER2 amplified patients with CDK12 fusions might also benefit from treatment with PARP1/2 inhibitors or platinum chemotherapy (13). To ascertain the frequency of CDK12 protein absence in HER2 amplified patients, and possible associations with outcome subsequent to anti-HER2 therapy, we assessed CDK12 protein expression in HER2 positive patients whom had been treated with Herceptin (21). Overall, 4/119 (3.4%) tumors were CDK12 absent and 71/119 (59.7%) tumors were CDK12 high (Fig. 1A). Lack of CDK12 expression did not improve survival in this cohort following Herceptin treatment ($p= 0.586$; HR= 22.636, 95% CI = 0-1718290, Log Rank (Mantel-Cox)), Fig. 2D).

In summary (Supplementary Table S1), overall, absence of CDK12 protein expression was seen in 10.5% of unselected BC; with a similar proportion of ER+ tumours showing absent CDK12 (9.8%). Interestingly, a higher frequency of absent CDK12 was seen in TNBC within this unselected cohort (17%); however, no association with BC specific survival was seen (p= 0.577, HR= 2, 95% CI = 1.8-3.4, Log Rank (Mantel-Cox)). As expected, in all three cohorts analysed, a lower proportion of HER2-positive tumours were CDK12 absent. Rather, most HER2 positive tumours were CDK12 high (protein and mRNA). Although high CDK12 expression was associated with a worse BC specific survival in these cohorts, this association was significant in the METABRIC dataset in multivariate analysis taking HER2 status into account, but not borne out in additional datasets.

Association of CDK12 expression and amplification

We subsequently investigated the associations of CDK12 mRNA expression with genomic status in primary BC from METABRIC (23) where there is copy number and gene expression data on the same tumor specimen. Of 1979 tumors with matched copy number and gene expression data available for CDK12, 208 (10.5%) harboured amplification encompassing the *CDK12* gene and a concurrent increase in its transcript expression, (p< 0.0001; Mann-Whitney U test; Fig. 3A). Of these, 99% (205/208) were also *HER2* amplified. There was a significant association with BC specific survival comparing *CDK12* amplified versus non-amplified tumors (p< 0.0001, Log Rank (Mantel-Cox), (HR = 0.45; 95% CI = 0.35 – 0.57), however, similarly to CDK12 protein expression this was lost in multivariate analysis taking HER2 status into account (Supplementary Table S4). Assessment of CDK12 protein expression by IHC in the METABRIC TMA, stratifying tumors as absent (0) versus highly expressed (7-8), however, revealed a significant association between *CDK12* amplification and increased CDK12 protein expression with 12/16 (75%) amplified

and 42/120 (35%) non-amplified tumors showing high protein expression, p= 0.0049, Fishers exact-test, Fig. 3B).

Recent evidence has pointed to the role of *CDK12* as a potential oncogene, given that that its amplification is associated with increased protein expression and aggressive clinical characteristics (19) and that *CDK12* amplified tumors show significantly increased CDK12 protein activation (39). To seek any evidence of oncogenic addiction to the downstream consequences of activated CDK12, we mined publicly available genetic perturbation screens in BC cell line models using validated reagents (12,13), to ascertain if *CDK12* amplified cells were addicted to CDK12 expression and downstream signalling for their survival. Analysis of both genome-wide pooled shRNA screen data (34) and siRNA kinome screen data (35) failed to identify any association between *CDK12* amplification and sensitivity to shRNA or siRNA designed to target *CDK12* (Fig. 3C-D**)**.

Genetic mechanisms of absent CDK12 expression

Given our previous findings that disruption of *CDK12* can occur as a result of an amplification breakpoint in the *ERBB2* amplicon that converges on *CDK12*, disrupting its expression (13), we sought to confirm our findings in primary tumors from METABRIC (23). Of all cases that showed *CDK12* amplification, 14.4% (30/208) harboured a breakpoint in *CDK12* that was associated with a significant reduction in *CDK12* transcript levels (p< 0.0001, Mann Whitney U test; Fig. 4A). In *HER2* amplified patients, there was no significant difference in BC specific survival between patients with a breakpoint in *CDK12* and those without (p= 0.32, Log Rank (Mantel-Cox) HR= 0.9897, 95% CI= 0.5107-1.918).

To ascertain the frequency of absent CDK12 protein in tumors with copy number alterations in *CDK12*, we intersected the IHC data from a subset of the METABRIC

cohort performed above with the available copy number data for *CDK12*. Of all *CDK12* amplified tumors, 14% (4/28) showed absent CDK12 protein expression and 33% (2/6) of tumors with a breakpoint in *CDK12* were CDK12 absent. This highlights that only a proportion of tumors with *CDK12* genomic breakpoints lead to loss of CDK12 protein expression.

In HGSOC*, CDK12* inactivating mutations have been reported to inactivate gene expression and consequently abrogate HR DNA repair pathways (10,11,14). Examination of DNA sequencing data from The Cancer Genome Atlas (TCGA) (32) identified *CDK12* mutations in 1.5% (20/1373) of unselected BC (Fig. 4B), (Supplementary Table S5 and S6)**.** These included 12 missense mutations, and 8 truncating mutations. Overall, 45% (9/20 (8 truncating and 1 missense) were predicted to disrupt protein function, and of these 25% (2/9) were seen in patients with triple-negative disease, whereas the remaining patients had ER+ or ER-/HER2+ disease. In contrast with recent data in HGSOC (11), predicted deleterious mutations did not correlate with reduced transcript expression, or with down-regulation of DNA damage response (DDR) genes, (Supplementary Table S7). CDK12 promoter methylation was a rare event seen in 1/696 (0.14%) of unselected BC (Supplementary Table S6). We next looked for other genomic alterations that would lead to absent CDK12 protein expression by interrogation of METABRIC cases with copy number and miRNA expression (23,40). This identified that heterozygous loss of CDK12 accounted for 7.1% (6/84) cases with absent CDK12 protein expression, however this was not enriched in the CDK12 absent group compared to the CDK12 high group ($p>0.999$, Fishers exact test). In addition, of the 14/162 miRNA's, that are known or predicted targets of CDK12, present in METABRIC (40) none showed correlation with CDK12 protein expression after multiple correction (Supplementary Table S8), suggesting there are additional mechanisms that lead to CDK12 protein loss.

Absent CDK12 protein expression is associated with reduced protein expression of genes involved in DNA repair

Although predicted deleterious mutations in *CDK12* did not correlate with reduced expression of DDR genes (see above), we assessed whether loss of CDK12 protein expression correlated with a reduction in the transcript expression of DNA repair proteins or biomarkers of DNA damage in the unselected series of BC cases (Table 1)**.** Although no significant correlations were observed (after multiple testing correction) with DDR genes at the mRNA level (Supplementary Table S9), absent CDK12 protein was significantly correlated with reduced protein expression of ATR, APE1, nuclear and cytoplasmic SMC6, Ku70/Ku80, DNA-PK and γH2AX nuclear positivity**.** Absent CDK12 protein expression also significantly correlated with both cleaved PARP1 (p= 0.003, Chi-Square) and non-cleaved PARP1 (p= 0.005, Chi-Square) expression suggesting that PARP1 levels are higher in tumors with absent CDK12. Interestingly, absent CDK12 expression also correlated with a decreased expression of TP53 (p= 0.001; Chi-Square) and RB1 (p= 0.003, Chi-Square), however there was no significant correlation between absent CDK12 protein and *TP53* mutations in the METABRIC cohort of tumors (17.5% CDK12 absent and 26.4% CDK12 high tumours harbouring *TP53* mutations p= 0.2779, Fishers exact test).

Since absent CDK12 protein expression was seen in 17% of TNBC within our analysis, we assessed this subset of tumors for correlations with the expression of DNA repair proteins as above. Even within this relative small subset, some significant correlations were still seen with DDR genes at the protein level: ATR (p= 0.018, Chi-Square); Ku70/Ku80 (p= 0.01, Chi-Square); and loss or decrease in nuclear and cytoplasmic SMC6 (p= 0.045, Chi-Square; Table 1). Of note comparison of CDK12 absent (0) and intermediate (2-6) levels of expression together as one group versus

high (7-8) failed to identify any significant associations with DNA repair genes both in unselected BC and TNBC (Supplementary Table S10). This suggests that functional loss of CDK12 is only observed in tumors with absent CDK12 protein expression in breast cancer and not low levels of CDK12 expression.

It has been shown that in HGSOC CDK12 mutations are consistently associated with a particular genomic instability pattern characterized by hundreds of tandem duplications of up to 10 megabases (Mb) in size, dubbed the 'CDK12 TD-plus phenotype' (14). Assessment of this pattern in both unselected and TNBC with absent CDK12 protein failed to identify any association with large numbers and sizes of tandem duplications (Fig. 4C-D). Together these results suggest that absent CDK12 protein in BC is associated with some defects in DNA repair related genes, however the resultant genomic scars are likely to be different to that seen in HGSOC (11).

Discussion

In this study, we were able to analyze, for the first time, the distribution of CDK12 protein in cohorts of primary BC. We show that high CDK12 expression is significantly correlated with HER2 status and that absent CDK12 is associated with a triple-negative phenotype and disruption of proteins involved in DNA repair.

The significant correlation of CDK12 with HER2 expression in these data is not surprising, since CDK12 is known to map to the smallest region within the HER2 amplicon (17,18). Therefore, although we observed significant correlations with CDK12 expression and patient survival, this significance was subsequently lost when HER2 positivity was taken into account. However, this association was substantiated in the METABRIC cohort at the protein level, suggesting investigation of additional

cohorts are warranted. It has been postulated that CDK12 may act in an oncogenic manner, given observed associations with amplification and increased transcript and protein expression and subsequent phosphorylation (19,39). This may perhaps occur through its reported roles in transcription through phosphorylation of RNA Pol II and regulation of pre-mRNA processing (5,7). By analyzing published siRNA and shRNA cell viability screens, using validated reagents (12,13) we show here that amplified cells are not addicted to CDK12 for their survival. We concede the assay length (average 4-7 days) in these experiments, may be too short to detect a loss of viability to depletion of a cyclin dependent kinase and additional time may be required to reduce DNA repair protein expression and accumulate damage required to affect viability. Interestingly, a recent study has suggested that CDK12 does act in an oncogenic manner in HER2 amplified cells by promoting cell invasion via alternative splicing and subsequent down-regulation of the long isoform of DNAJB6 (41), suggestive of a mechanism by which CDK12 acts in an oncogenic manner to promote cell invasiveness. Of note, elevated levels of CDK12 in HER2 positive tumors have been a proposed reason why tumors arising in *BRCA1* carriers are usually ERBB2-negative, given elevated expression would oppose the defects in HR mediated by *BRCA1* loss (34).

We have shown previously that a proportion of HER2-positive tumors harbor inactivating (out-of-frame) fusion genes in *CDK12* that are due to a copy number breakpoint in the HER2 amplicon converging on *CDK12*, resulting in a significant decrease in both transcript and protein levels (13). Furthermore, cell lines with breakpoints in *CDK12* that result in loss of protein expression are sensitive to PARP1/2 inhibitor therapy, due to impaired HR mediated DNA repair (13). By assessing the distribution of CDK12 protein expression in CDK12 amplified breast cancers in the METABRIC cohort, we identified 14% that were CDK12 absent. Moreover, of the tumors with a breakpoint in CDK12, 33% had absent CDK12 protein

expression. Whilst the numbers we were able to assess are small, this nevertheless suggests a proportion of HER2 positive patients show absent CDK12 protein as a result of both copy number breakpoints and additional mechanisms. Overall, our data suggest there may be a small proportion of HER2-positive patients that may benefit from treatment with DNA damage response targeted therapies such as PARP1/2 inhibitor therapy.

Recent data in HGSOC has shown that *CDK12* point mutations inactivate gene expression subsequently abrogating HR DNA repair pathways (10,11). Although in our study *CDK12* mutations did not themselves correlate with decreased expression of DNA repair genes (perhaps due to residual CDK12 activity, given some of these are not associated with loss of heterozygosity), we show that BC with absent CDK12 protein show down-regulation of a number of genes involved in functional HR DNA repair at the protein level, suggesting that absent CDK12 protein in BC may also be associated with HR defects. Although we were unable to assess all proteins known to be dysregulated in CDK12 mutant HGSOC such as RAD51, FANCI and FANCD2, it is reasonable to postulate that CDK12 may have other targets, including DNA repair proteins themselves, phosphorylation of which is required for protein stability. Lack of any observed associations at the mRNA level, may be due to small numbers in the study, however, our observed associations warrant further investigation in larger cohorts. Recent evidence in HGSOC points to a unique signature of genomic instability in *CDK12* mutated tumors that is indicative of gross defects in DNA repair characterized by tens to hundreds of large tandem duplications scattered throughout the genome, although this was not observed in *CDK12* mutated breast cancers (14). Consistent with this, we observed no correlation with large tandem duplications of neither all BC nor TNBC specifically. It is intriguing however that we observed absent CDK12 protein in 17% of TNBC from the unselected primary BC analyzed in this study. Of course, this could be a consequence of the small number of cases

included, however CDK12 inactivation may play a role in a subset of TNBC's possibly through mechanisms distinct from those observed in HGSOC.

In conclusion, we have shown that a subset of HER2-positive patients show absent CDK12 protein expression and have shown an enrichment of absent CDK12 protein expression in TNBC. Moreover, we have provided evidence that absent CDK12 expression is associated with defects in DNA repair proteins. These results suggest that CDK12 IHC could potentially be useful, once validated in sufficiently powered studies, for stratification of patients for treatment with DDR targeted therapies.

Author's contributions:

Conception and design: K. Naidoo**,** C.J. Lord and R. Natrajan

Development of methodology: K. Naidoo, F. Daley and R. Natrajan

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. Naidoo, P. Wai, F. Daley, D. Kriplani, A. Tutt, S. Madhusudan E. Rhaka, E, I.O. Ellis, S.Y. Chan, P.M. Moseley, T.M.A. Abdel-Fatah and A.R. Green,

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): K. Naidoo, S. Maguire, S. Haider, J. Campbell, H. Mirza, A. Grigoriadis, Y. Yuan, A.R. Green and R. Natrajan.

Writing, review, and/or revision of the manuscript: K. Naidoo, P. Wai, F. Daley, J. Campbell, A. Grigoriadis, A. Tutt, S.Y. Chan, P.M. Moseley, E. Rakha, I.O. Ellis, C.J. Lord, Y. Yuan, A.R. Green and R. Natrajan

Acknowledgements:

The results shown here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. This study makes use of

METABRIC data generated by the Molecular Taxonomy of Breast Cancer International Consortium with funding provided by Cancer Research UK and the British Columbia Cancer Agency Branch. We acknowledge NHS funding to the ICR/Royal Marsden Hospital Biomedical Research Centre. We thank J Watkins for providing the tandem duplication R script.

Grant support:

HM, AG and AT are supported by the Breast Cancer Now Research Unit (former Breakthrough Breast Cancer Research) funding at King's College London. We also thank the Breast Cancer Now Tissue Bank/Nottingham Health Science Biobank (NHSB) for tissue provision.

References

1. Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet **2012**;379(9814):432-44.
2. Early Breast Cancer Trialists' Collaborative G, Davies C, Godwin J, Gray R, Clarke M, Cutter D, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet **2011**;378(9793):771-84.
3. Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. Lancet Oncol **2010**;11(12):1135-41.
4. Gianni L, Dafni U, Gelber RD, Azambuja E, Muehlbauer S, Goldhirsch A, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. Lancet Oncol **2011**;12(3):236-44.
5. Cheng SW, Kuzyk MA, Moradian A, Ichu TA, Chang VC, Tien JF, et al. Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. Mol Cell Biol **2012**;32(22):4691-704.
6. Chen HH, Wang YC, Fann MJ. Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. Mol Cell Biol **2006**;26(7):2736-45.
7. Liang K, Gao X, Gilmore JM, Florens L, Washburn MP, Smith E, et al. Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13

complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. Mol Cell Biol **2015**;35(6):928-38.
8. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev **2011**;25(20):2158-72.
9. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature **2011**;474(7353):609-15.
10. Joshi PM, Sutor SL, Huntoon CJ, Karnitz LM. Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. J Biol Chem **2014**;289(13):9247-53.
11. Ekumi KM, Paculova H, Lenasi T, Pospichalova V, Bosken CA, Rybarikova J, et al. Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. Nucleic Acids Res **2015**;43(5):2575-89.
12. Bajrami I, Frankum JR, Konde A, Miller RE, Rehman FL, Brough R, et al. Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. Cancer Res **2014**;74(1):287-97.
13. Natrajan R, Wilkerson PM, Marchio C, Piscuoglio S, Ng CK, Wai P, et al. Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. J Pathol **2014**;232(5):553-65.
14. Popova T, Manie E, Boeva V, Battistella A, Goundiam O, Smith NK, et al. Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. Cancer Res 2016; 76(7):1882-91
15. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature **2012**;487(7406):239-43.
16. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell **2015**;161(5):1215-28.
17. Arriola E, Marchio C, Tan DS, Drury SC, Lambros MB, Natrajan R, et al. Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines. Lab Invest **2008**;88(5):491-503.
18. Sircoulomb F, Bekhouche I, Finetti P, Adelaide J, Ben Hamida A, Bonansea J, et al. Genome profiling of ERBB2-amplified breast cancers. BMC Cancer **2010**;10:539.
19. Capra M, Nuciforo PG, Confalonieri S, Quarto M, Bianchi M, Nebuloni M, et al. Frequent alterations in the expression of serine/threonine kinases in human cancers. Cancer Res **2006**;66(16):8147-54.
20. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology **2002**;41(3A):154-61.
21. Green AR, Barros FF, Abdel-Fatah TM, Moseley P, Nolan CC, Durham AC, et al. HER2/HER3 heterodimers and p21 expression are capable of predicting adjuvant trastuzumab response in HER2+ breast cancer. Breast Cancer Res Treat **2014**;145(1):33-44.
22. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer **2005**;116(3):340-50.

23. 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(7403):346-52.
24. Green AR, Aleskandarany MA, Agarwal D, Elsheikh S, Nolan CC, Diez-Rodriguez M, et al. MYC functions are specific in biological subtypes of breast cancer and confers resistance to endocrine therapy in luminal tumours. Br J Cancer **2016**;114(8):917-28.
25. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer **2007**;109(1):25-32.
26. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, et al. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clin Cancer Res **2009**;15(7):2302-10.
27. Aleskandarany MA, Green AR, Benhasouna AA, Barros FF, Neal K, Reis-Filho JS, et al. Prognostic value of proliferation assay in the luminal, HER2-positive, and triple-negative biologic classes of breast cancer. Breast Cancer Res **2012**;14(1):R3.
28. Aleskandarany MA, Green AR, Rakha EA, Mohammed RA, Elsheikh SE, Powe DG, et al. Growth fraction as a predictor of response to chemotherapy in node-negative breast cancer. Int J Cancer **2010**;126(7):1761-9.
29. Aleskandarany MA, Rakha EA, Ahmed MA, Powe DG, Ellis IO, Green AR. Clinicopathologic and molecular significance of phospho-Akt expression in early invasive breast cancer. Breast Cancer Res Treat **2011**;127(2):407-16.
30. Aleskandarany MA, Rakha EA, Ahmed MA, Powe DG, Paish EC, Macmillan RD, et al. PIK3CA expression in invasive breast cancer: a biomarker of poor prognosis. Breast Cancer Res Treat **2010**;122(1):45-53.
31. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature **2012**;490(7418):61-70.
32. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell **2015**;163(2):506-19.
33. Todd JR, Ryall KA, Vyse S, Wong JP, Natrajan RC, Yuan Y, et al. Systematic analysis of tumour cell-extracellular matrix adhesion identifies independent prognostic factors in breast cancer. Oncotarget **2016**;7(39):62939-53.
34. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, et al. Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. Cell **2016**;164(1-2):293-309.
35. Campbell J, Ryan CJ, Brough R, Bajrami I, Pemberton HN, Chong IY, et al. Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines. Cell Rep **2016**;14(10):2490-501.
36. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A **2010**;107(39):16910-5.
37. Watkins J, Tutt A, Grigoriadis A. Tandem duplications contribute to not one but two distinct phenotypes. Proc Natl Acad Sci U S A **2016**;113(36):E5257-8.
38. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat **2006**;100(2):229-35.
39. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature **2016**;534(7605):55-62.
40. Dvinge H, Git A, Graf S, Salmon-Divon M, Curtis C, Sottoriva A, et al. The shaping and functional consequences of the microRNA landscape in breast cancer. Nature **2013**;497(7449):378-82.

41. Tien JF, Mazloomian A, Cheng SG, Hughes CS, Chow CC, Canapi LT, et al. CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. Nucleic Acids Res 2017. doi: 10.1093/nar/gkx187
42. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal **2013**;6(269):pl1.
43. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov **2012**;2(5):401-4.

Figure Legends

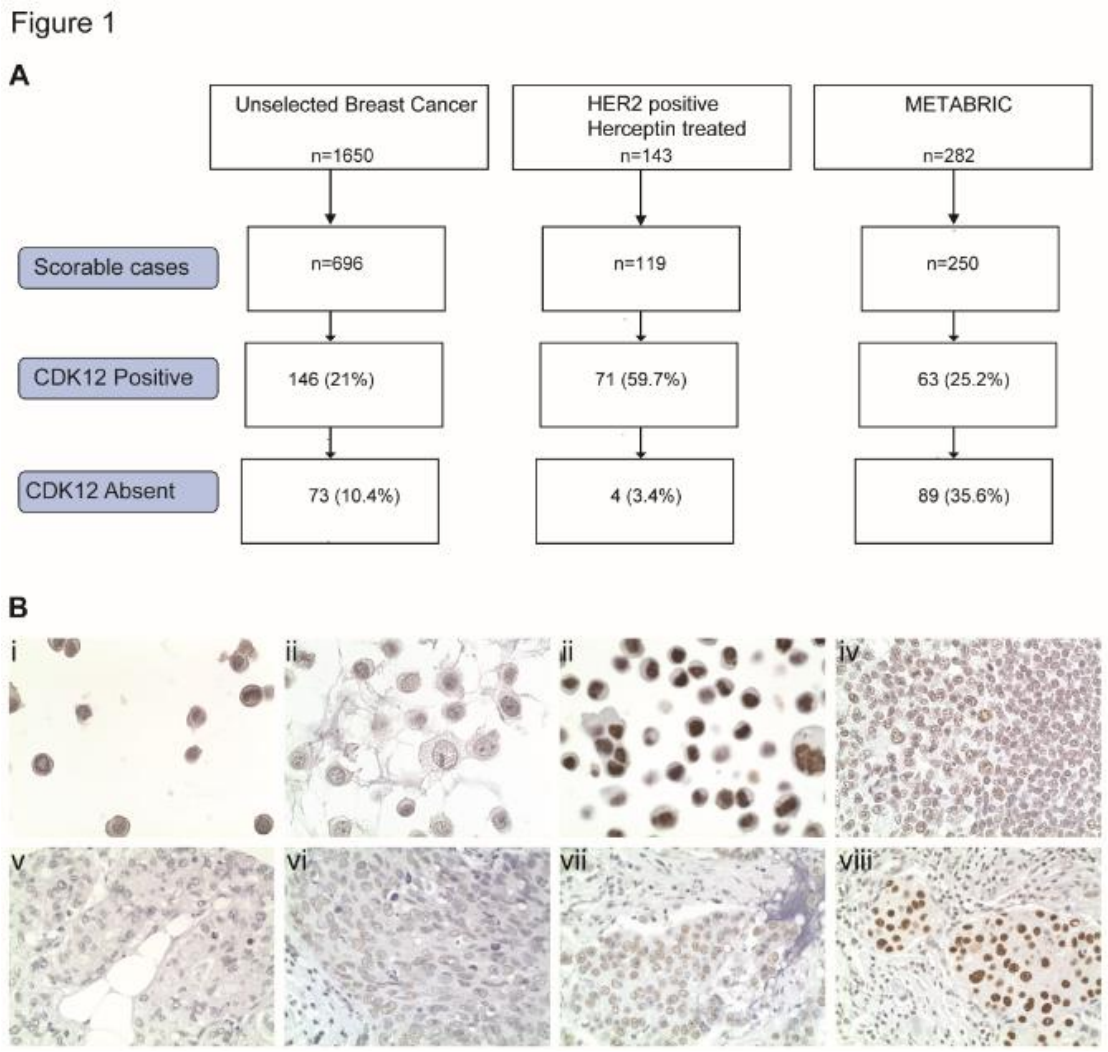

Figure 1: Distribution of CDK12 protein expression in breast cancer

A, Modified CONSORT diagram depicting the distribution of CDK12 positive and negative breast cancers in each of the cohorts analysed. **B,** Representative micrographs of CDK12 protein expression in **i,** MCF7 cell line treated with non-

targeting siRNA controls; **ii,** MCF7 cell line treated with previously validated siRNA against CDK12 (13); **iii.** BT474 CDK12 amplified cells all at x400 magnification; **iv,** tonsil positive control (x 200 magnification). **v-viii,** Representative images of staining intensity in primary breast cancers, where CDK12 expression was quantified using a modified Allred score, which assessed both intensity (highest score = 3) and percentage positivity (highest score = 5): **(v)** negative; **(vi)** 1+; (**vii)** 2+; and **(viii)** 3+; all images at 200x magnification. A score of 0 was considered absent and a score of 7 or 8, as high expression.

Figure 2

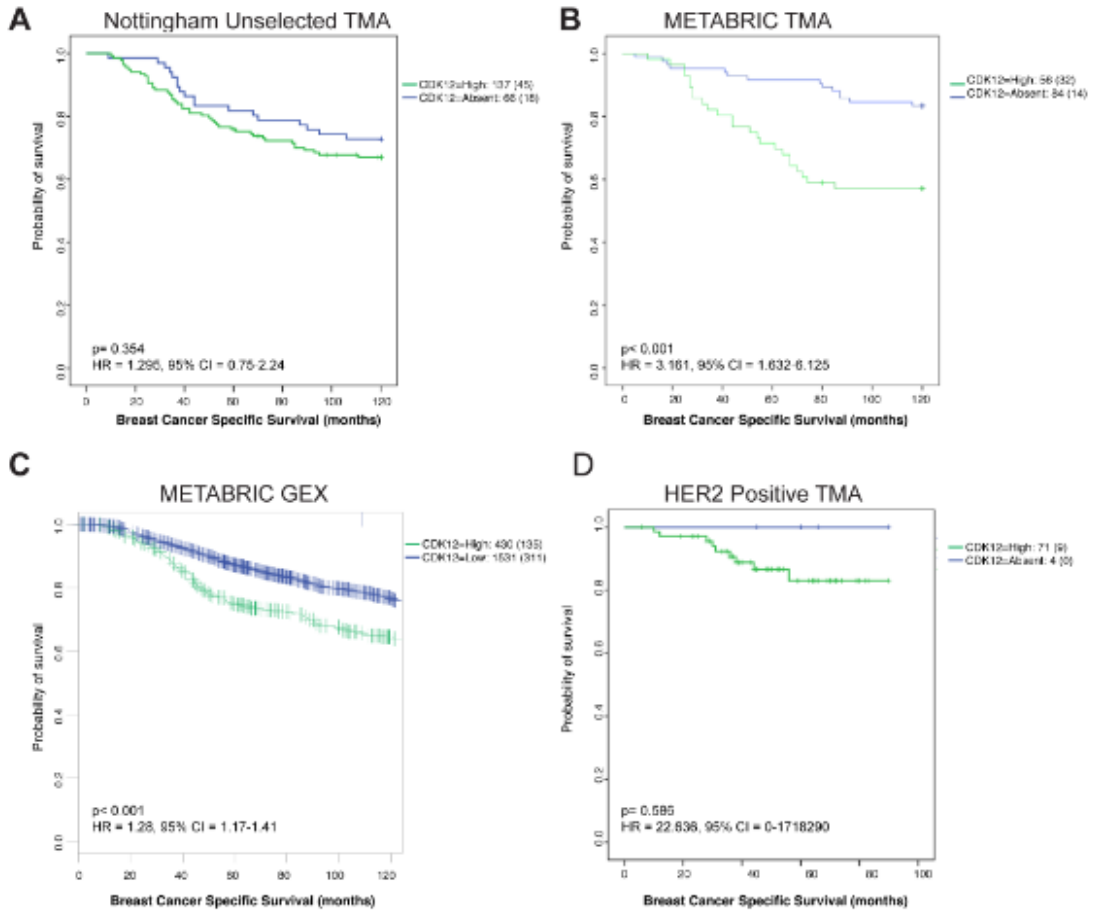

Figure 2: CDK12 positive breast cancer has a poorer survival in univariate analysis

Kaplan-Meier curves showing breast cancer specific survival for CDK12 high (7-8) versus negative (0) breast cancers asses by IHC in **A,** Nottingham unselected

primary breast cancer series (n= 203); **B,** tumors from METABRIC (n= 140); **C,** Gene expression correlations of CDK12 low versus high from METABRIC (n= 1961) and **D,** HER2-positive tumors treated with Herceptin (n= 75).

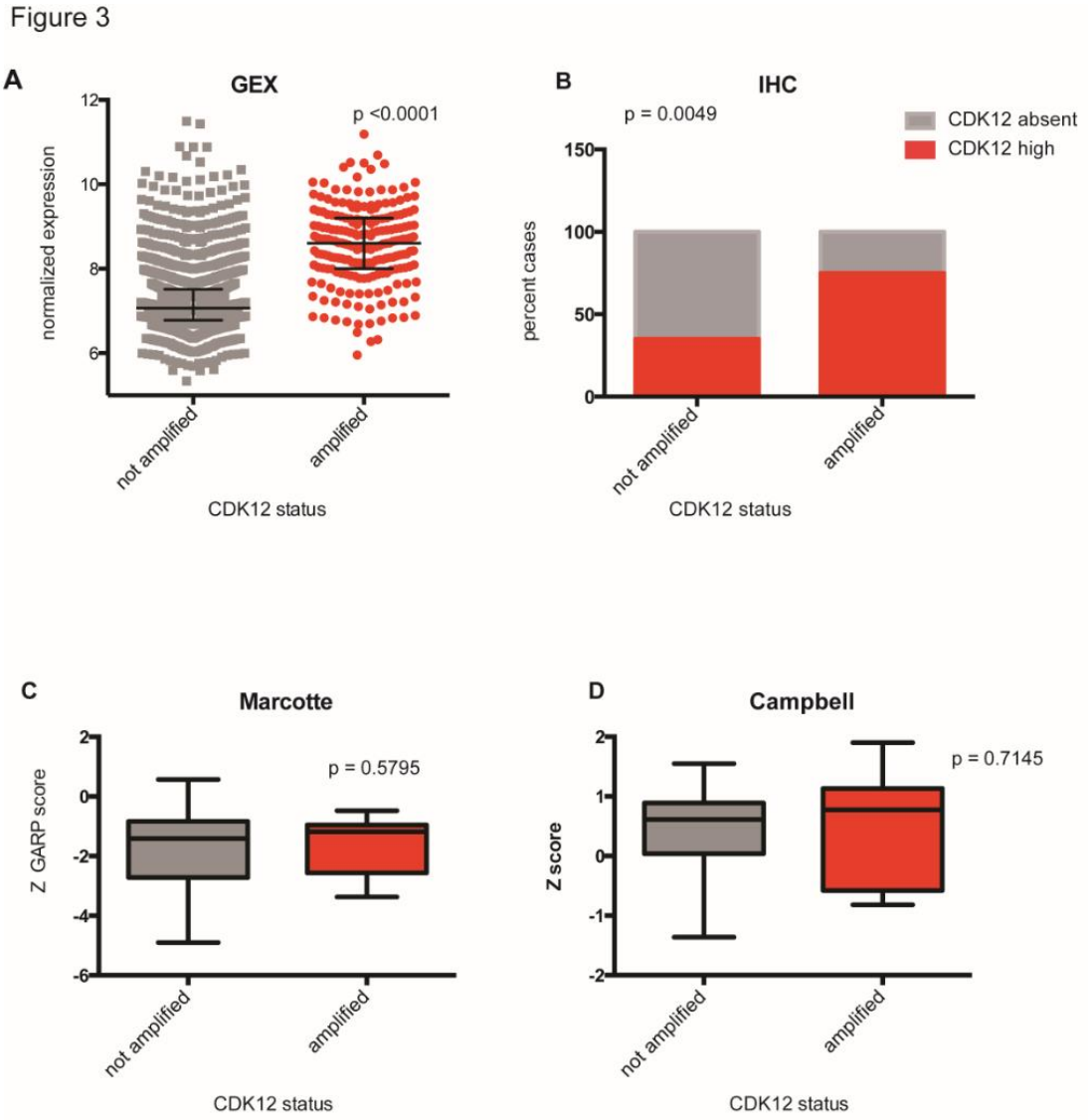

Figure 3: CDK12 amplified cells are not dependent on CDK12 expression for their survival

A, Scatter dot plot depicting a significant association of *CDK12* transcript expression with gene amplification (n= 208) versus no amplification (n= 1769) (error bars represent median with the interquartile range). **B,** bar-chart depicting a significant increase of CDK12 protein expression as measured by IHC in *CDK12* amplified tumors (n= 16) compared to non-amplified (n= 103). **C-D,** Relative cell viability after

CDK12 silencing in *CDK12* amplified (Campbell, n= 8; Marcotte n= 14) versus non-amplified cell lines (Campbell, n= 19; Marcotte n= 40), showing no significant difference in cell survival from c) Marcotte et al (34) and d) Campbell at al (35).

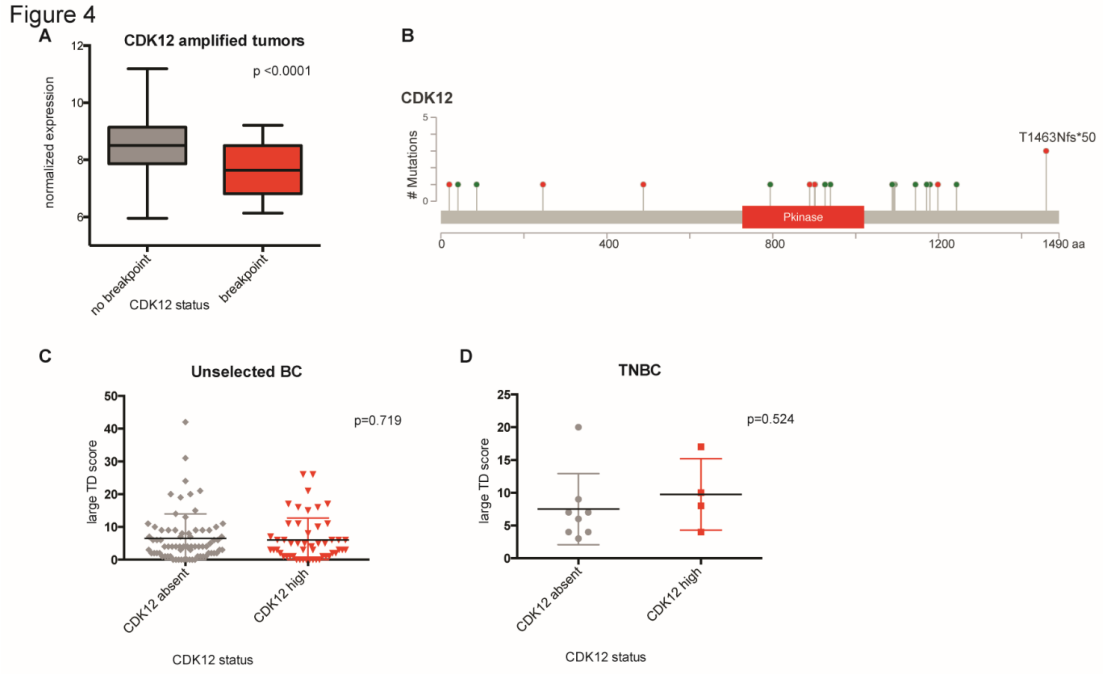

Figure 4: CDK12 protein loss is associated with DNA repair defects in breast cancer

A, Box and whisker plot (min-max) showing a significant decrease in *CDK12* transcript levels in HER2-amplified tumors with breakpoints (n= 30) in *CDK12* compared with no breakpoints (n= 178) from METABRIC. **B,** Lollipop diagram depicting the distribution of *CDK12* mutations in breast cancer (red= frameshift and nonsense mutations, green= non-synonymous coding mutations). **C,** Scatter dot plot diagrams showing significant associations between CDK12 absent (n= 74) (IHC score 0) versus CDK12 high (n= 51) (IHC score 7-8) in unselected METABRIC tumors with large tandem duplication score, indicative of gross genomic defects. **D,** Scatter dot plot diagrams showing significant associations between CDK12 absent (n= 8) (IHC score 0) versus CDK12 high (n= 4) (IHC score 7-8) in TNBC from

METABRIC, with large tandem duplication score, indicative of gross genomic defects.

Table 1: Association of CDK12 expression with DNA repair proteins in unselected and TNBC

	CDK12 Absent (0) *No. of cases (%)*	**CDK12 High (7/8)** *No. of cases (%)*	**p-value**
Unselected Series			
BRCA1 (n= 170)			0.249
Negative	27 (51.9)	50 (42.4)	
Positive	25 (48.1)	68 (57.6)	
PARP1 cleaved (n= 170)			0.003**
Negative	18 (30)	13 (11.8)	
Positive	42 (70)	97 (88.2)	
PARP1 non-cleaved (n= 175)			0.005**
Negative	36 (62.1)	46 (39.3)	
Positive	22 (37.9)	71 (60.7)	
ATR (n= 173)			<0.001**
Negative	46 (70.8)	45 (41.7)	
Positive	19 (29.2)	63 (58.3)	
APE1 (n= 89)			0.02*
Negative	10 (38.5)	10 (15.9)	
Positive	16 (61.5)	53 (84.1)	
Ku70/Ku80 (n= 153)			0.001**
Negative	18 (34.0)	11 (11.0)	
Positive	35 (66.0)	89 (89.0)	
DNA-PK (n= 156)			0.002**
Negative	17 (32.1)	12 (11.7)	
Positive	36 (67.9)	91 (88.3)	
SMC6 cytoplasmic (n= 166)			0.002**
Negative	30 (54.5)	33 (29.7)	
Positive	25 (45.5)	78 (70.3)	
SMC6 nuclear (n= 166)			< 0.001**
Negative	33 (60.0)	34 (30.6)	
Positive	22 (40.0)	77 (69.4)	
γH2AX (n= 150)			0.002**
Negative	24 (47.1)	22 (22.2)	
Positive	27 (52.9)	77 (77.8)	
TP53 (n= 209)			0.001**
Negative	56 (80.0)	80 (57.6)	
Positive	14 (20.0)	59 (42.4)	
CHEK1 cytoplasmic (n= 103)			0.321
Negative	29 (50.9)	44 (42.7)	
Positive	28 (49.1)	59 (57.3)	
CHEK1 nuclear (n= 161)			0.927
Negative	48 (84.2)	87 (83.7)	
Positive	9 (15.8)	17 (16.3)	
CHEK2 (n= 108)			0.345
Negative	31 (50.8)	63 (58.3)	
Positive	30 (49.2)	45 (41.7)	

Triple Negative Breast Cancer			
BRCA1 (n= 28)			0.430
Negative	10 (71.4)	8 (57.1)	
Positive	4 (28.6)	6 (42.9)	
PARP1 cleaved (n= 28)			0.378
Negative	8 (50.0)	4 (33.3)	
Positive	8 (50.0)	8 (66.7)	
PARP1 non-cleaved (n= 29)			0.103
Negative	11 (68.8)	5 (38.5)	
Positive	5 (31.2)	8 (61.5)	
ATR (n= 33)			0.018*
Negative	18 (90.0)	7 (53.8)	
Positive	2 (10.0)	6 (46.2)	
APE1 (n= 17)			0.453
Negative	2 (25.0)	1 (11.1)	
Positive	6 (75.0)	8 (88.9)	
Ku70/Ku80 (n= 27)			0.010*
Negative	8 (47.1)	0 (0.0)	
Positive	9 (52.9)	10 (100.0)	
DNA-PK (n= 31)			0.283
Negative	6 (37.5)	3 (25.0)	
Positive	10 (62.5)	12 (75.0)	
SMC6 cytoplasmic (n= 31)			0.018*
Negative	9 (60.0)	3 (23.1)	
Positive	6 (40.0)	13 (76.9)	
SMC6 nuclear (n= 31)			0.045*
Negative	11 (73.3)	6 (37.5)	
Positive	4 (26.7)	10 (62.5)	
γH2AX (n= 29)			0.089
Negative	9 (60.0)	4 (28.6)	
Positive	6 (40.0)	10 (71.4)	
TP53 (n= 36)			0.187
Negative	12 (63.2)	7 (41.2)	
Positive	7 (36.8)	10 (58.8)	
CHEK1 cytoplasmic (n= 27)			0.384
Negative	8 (47.1)	3 (30.0)	
Positive	9 (52.9)	7 (70.0)	
CHEK1 nuclear (n= 28)			0.068
Negative	17 (100.0)	9 (81.8)	
Positive	0 (0.0)	2 (18.2)	
CHEK2 (n= 32)			0.961
Negative	13 (68.4)	9 (69.2)	
Positive	6 (31.6)	4 (30.8)	

Supplementary Tables:

Supplementary Table S1: CDK12 expression in relation to clinicopathological parameters for the unselected TMA series

Supplementary Table S2: CDK12 expression in relation to clinicopathological parameters for the HER2-positive Herceptin treated series

Supplementary Table S3: CDK12 expression in relation to clinicopathological parameters for the METABRIC TMA series

Supplementary Table S4: Univariate and multivariate analysis of CDK12 in the TMA cohorts

Supplementary Table S5: CDK12 mutations in breast cancer. Taken from cBioportal (42,43).

Supplementary Table S6: Correlations of CDK12 mutations, methylation, gene expression and ERBB2 copy number in primary breast cancers from TCGA

Supplementary Table S7: Correlations of CDK12 mutations and gene expression of DNA repair genes in primary tumors from METABRIC. P values from heteroscedastic 2-tailed, t-test

Supplementary Table S8: Correlations of CDK12 protein expression, and miRNA expression in primary tumors from METABRIC. Wilcoxon rank P values are corrected for multiple testing.

Supplementary Table S9: Correlations of CDK12 protein expression and gene expression of DNA repair genes in primary tumors from METABRIC. Limma analysis corrected for multiple testing.

Supplementary Table S10: Association of CDK12 absent and intermediate (0, 2-6) versus high (7-8) expression with DNA repair proteins in unselected and TNBC. P values from Fishers exact test.