

RESEARCH ARTICLE

Open Access

# Systematic antibody generation and validation via tissue microarray technology leading to identification of a novel protein prognostic panel in breast cancer

Patrick C O'Leary<sup>1†</sup>, Sarah A Penny<sup>1†</sup>, Roisin T Dolan<sup>1,6†</sup>, Catherine M Kelly<sup>1†</sup>, Stephen F Madden<sup>3</sup>, Elton Rexhepaj<sup>1</sup>, Donal J Brennan<sup>1</sup>, Amanda H McCann<sup>1,2</sup>, Fredrik Pontén<sup>4</sup>, Mathias Uhlén<sup>5</sup>, Radoslaw Zagodzón<sup>1</sup>, Michael J Duffy<sup>2,7</sup>, Malcolm R Kell<sup>6</sup>, Karin Jirstrom<sup>8†</sup> and William M Gallagher<sup>1\*†</sup>

## Abstract

**Background:** Although omic-based discovery approaches can provide powerful tools for biomarker identification, several reservations have been raised regarding the clinical applicability of gene expression studies, such as their prohibitive cost. However, the limited availability of antibodies is a key barrier to the development of a lower cost alternative, namely a discrete collection of immunohistochemistry (IHC)-based biomarkers. The aim of this study was to use a systematic approach to generate and screen affinity-purified, mono-specific antibodies targeting progression-related biomarkers, with a view towards developing a clinically applicable IHC-based prognostic biomarker panel for breast cancer.

**Methods:** We examined both in-house and publicly available breast cancer DNA microarray datasets relating to invasion and metastasis, thus identifying a cohort of candidate progression-associated biomarkers. Of these, 18 antibodies were released for extended analysis. Validated antibodies were screened against a tissue microarray (TMA) constructed from a cohort of consecutive breast cancer cases ( $n = 512$ ) to test the immunohistochemical surrogate signature.

**Results:** Antibody screening revealed 3 candidate prognostic markers: the cell cycle regulator, Anillin (ANLN); the mitogen-activated protein kinase, PDZ-Binding Kinase (PBK); and the estrogen response gene, PDZ-Domain Containing 1 (PDZK1). Increased expression of ANLN and PBK was associated with poor prognosis, whilst increased expression of PDZK1 was associated with good prognosis. A 3-marker signature comprised of high PBK, high ANLN and low PDZK1 expression was associated with decreased recurrence-free survival ( $p < 0.001$ ) and breast cancer-specific survival (BCSS) ( $p < 0.001$ ). This novel signature was associated with high tumour grade ( $p < 0.001$ ), positive nodal status ( $p = 0.029$ ), ER-negativity ( $p = 0.006$ ), Her2-positivity ( $p = 0.036$ ) and high Ki67 status ( $p < 0.001$ ). However, multivariate Cox regression demonstrated that the signature was not a significant predictor of BCSS (HR = 6.38; 95% CI = 0.79-51.26,  $p = 0.082$ ).

**Conclusions:** We have developed a comprehensive biomarker pathway that extends from discovery through to validation on a TMA platform. This proof-of-concept study has resulted in the identification of a novel 3-protein prognostic panel. Additional biochemical markers, interrogated using this high-throughput platform, may further augment the prognostic accuracy of this panel to a point that may allow implementation into routine clinical practice.

**Keywords:** Prognostic biomarkers, Tissue microarray, Breast cancer, Antibody screening, Antibody validation

\* Correspondence: william.gallagher@ucd.ie

†Equal contributors

<sup>1</sup>UCD School of Biomolecular and Biomedical Science, UCD Conway Institute of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland

Full list of author information is available at the end of the article

## Background

Breast cancer is a heterogeneous disease driven by a continuum of mutations and abnormal gene/protein expression that controls the tumourigenic phenotype and molecular mechanisms underpinning the complexity of its clinical behaviour [1]. To select systemic therapies, current treatment guidelines combine traditional prognostic factors (stage, tumour size, histologic grade, nodal status) with estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (Her2) expression status. However, these conventional prognostic algorithms are insufficient to capture the biologic diversity of breast cancer and impede effective tailoring of individualised treatment strategies [2]. In the post-genomic era, advances in prognostic and predictive models are beginning to capture this heterogeneity, not least with the recent generation of a new molecular classification consisting of at least ten different breast cancer subtypes [3-6]. Molecular profiling of cancer tissues has aided the development of targeted therapies, improved our understanding of treatment resistance, and helps better predict patient prognosis. This knowledge has allowed personalised breast cancer therapeutic regimens to become an achievable goal.

The cornerstone of molecular profiling has historically been transcriptomics which has transformed our understanding of the complexity of the underlying signalling pathways and interactions within a breast tumour, as well as allowing the identification of gene expression signatures associated with patient outcome [4,7]. Consequently, clinical development of transcriptomic profiling tools has dramatically escalated, augmenting standard diagnostic and prognostic information obtained from traditional clinicopathological variables [8]. The most clinically advanced prognostic gene expression signatures in breast cancer are MammaPrint [7,9] and OncotypeDx [10], which are currently the subject of large-scale prospective randomised control trials to assess their utility for stratification of breast cancer patients [11-13].

Whilst transcriptomic approaches have undoubtedly enabled the acceleration of translational pathology, providing an excellent platform for omic-based discovery [13,14], reservations have been raised regarding the clinical applicability of gene expression studies given their prohibitive cost, often reliance on frozen tissue, quality assurance issues and the advanced technical expertise required to utilise the technology [2]. Crucially, mRNA transcription does not necessarily translate to protein expression, and it is not uncommon to observe a discrepancy between mRNA and protein expression [15,16]. As proteins are one of the primary effectors of the cell, protein-based assays may be more clinically relevant as biomarkers in personalised medicine. Effective implementation of personalised cancer therapy depends upon the

successful identification and translation of informative biomarkers to aid treatment provision. In a prior review, we described the contribution of antibody-based proteomics for fast-tracking the development of new diagnostic assays that are crucial to achieving personalisation of cancer therapy [17]. The systematic generation and validation of specific antibodies offers a high-throughput mechanism for the functional exploration of the proteome and a logical approach for fast-tracking the translation of identified biomarkers [17]. Whilst DNA microarray technology provides an excellent platform for biomarker discovery, it would now appear that IHC and genomic sequencing may play an increasingly important role in the clinical management of breast cancer [2]. Tissue microarrays (TMAs) are an ideal platform for rapid development of an IHC profile, allowing multiple targets to be systematically assessed, and reduce an assay to clinical utility [3-5,8,18-23].

In this proof-of-concept study, we used a novel high-throughput system, using affinity-purified, mono-specific antibodies, to translate protein targets from gene expression studies into clinically applicable IHC-based prognostic panels for breast cancer.

## Methods

### Selection of candidate biomarkers from transcriptomic datasets

Thirty-one genes were selected from an in-house analysis of the van 't Veer study [7], using a Between Group Analysis (BGA) method identifying the top 100 good and poor prognosis genes [24,25]. From this list, we considered the top 15 genes associated with good prognosis and the top 16 genes associated with poor prognosis. Another 25 genes of interest were selected from a transcriptomic study of ductal carcinoma *in situ* (DCIS) to invasive ductal carcinoma (IDC) progression, with a particular focus on transcripts that were up-regulated in the invasive component [26] (Additional file 1: Table S1).

### Patients

The TMAs used in this study were derived from a reference cohort of 512 consecutive invasive breast cancer cases diagnosed at the Department of Pathology, Malmö University Hospital, Malmö, Sweden between 1988 and 1992 and have been previously described [27-29]. The median patient age was 65 years (range 27-96) and median follow-up time regarding disease-specific and overall survival was 11 years (range 0-17). Duplicate cores for each patient were reported as consensus scores. Each patient was assigned a unique identifier that was then linked to an anonymised ethics board-approved database containing follow-up information. Patients with recurrent disease and previous systemic therapies were excluded. Two hundred and sixty-three patients were

deceased at the last follow-up date (December 2004), 90 of which were classified as breast cancer-specific deaths. Ethical permission was obtained from the Local Ethics Committee at Lund University (Dnr 613/02), whereby informed consent was deemed not to be required, but opting out was an option.

#### **TMA construction**

The TMAs were constructed using a manual tissue arrayer (MTA-1, Beecher Inc., WI, USA). PBK and PDZK1 were screened on a TMA inclusive of all 512 cases from the reference cohort with 0.6 mm duplicate tissue cores extracted from each donor block. ANLN was screened on a second generation TMA inclusive of 498 cases from the reference cohort, with 1.0 mm duplicate tissue cores extracted from each donor block and transferred to the recipient block. The total number of cores per block was limited to ~ 200 (100 patients), with a total of 5 blocks arrayed.

#### **Antibody generation**

The Human Protein Atlas (HPA) [30] use a high-throughput method to generate affinity-purified, mono-specific antibodies raised to all non-redundant human proteins [31]. Protein epitope sequence tag (PrEST)-specific antibodies represent unique regions of each protein target. Rabbit polyclonal antisera immunised with His<sub>6</sub>ABP-PrEST antigens derived from a subset of the 56 targets of interest described above (Additional file 1: Table S1) were purified by a two-step immunoaffinity protocol to obtain pure mono-specific antibodies [32].

#### **Cell culture**

A panel of breast epithelial cell lines were selected to test antibody specificity, including MCF-7, BT474, T47D, SKBR3, MDA-MB-231 and Hs578T cells. The Hs578T (i8) invasive subclone was a kind gift from Dr. Susan McDonnell (School of Chemical & Bioprocess Engineering, University College Dublin, Ireland) and was derived from the parental Hs578T cell line (also denoted as Hs578T(P)) by sequential selection through the BD Matrigel<sup>®</sup> Invasion Chamber assay system [33]. All remaining cell lines were purchased from the European Collection of Cell Cultures (Wiltshire, UK). The MCF-7, BT474, T47D, SKBR3, and MDA-MB-231 cell lines were cultured in DMEM supplemented with 10% (w/v) foetal calf serum, 2 mM L-glutamine, 50 IU/ml penicillin, and 50 µg/ml streptomycin sulphate. The Hs578T variants were also supplemented with 10 µg/ml bovine insulin. Cells were maintained in humidified air with 5% CO<sub>2</sub> at 37°C. Studies of protein expression were performed on cells at 70-80% confluence. All cell lines were routinely screened for Mycoplasma contamination.

#### **Western blot analysis**

Total protein was extracted from sub-confluent cells by the addition of radioimmunoprecipitation assay buffer (RIPA), followed by centrifugation at 16,000 g for 20 min at 4°C. The supernatants were removed and the protein levels determined using the bicinchoninic acid (BCA) method (Pierce, IL). Samples containing 50 µg aliquots of protein were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), on a 12% polyacrylamide gel under reducing conditions. Following electrophoresis, proteins were transferred to polyvinylidene fluoride membrane. Membranes were blocked in 5% non-fat milk for 1 hr at room temperature. Protein expression was detected using rabbit mono-specific polyclonal anti-human antibodies (HPA, Sweden) applied overnight at 4°C (PDZK1 1:1000 dilution; PBK, ANLN 1:500). Membranes were washed in TBS-T (Tris buffered saline with 0.1% Tween 20) and incubated for 1 hr with horseradish peroxidase (HRP)-conjugated anti-rabbit immunoglobulin (all antibodies: 1:5000 dilution). The blots were again washed in TBS-T. HRP was detected using Enhanced Chemiluminescence plus (Amersham Biosciences, UK). Chemiluminescence was detected by autoradiography using X-ray film. Membranes were stripped and re-probed with anti-β-actin (1:5000 dilution; Abcam, UK) as a loading control.

#### **Cell pellet arrays**

In order to validate the Western blotting results in the IHC setting, a cell pellet array was constructed and IHC was performed on the same panel of breast cancer cell lines. Cells were trypsinised and fixed for 1 hr in 10% formalin, centrifuged at 500 x g for 10 minutes, washed twice with PBS and re-suspended in 0.8% agarose. The tumour cell-containing agarose plugs were processed through gradient concentrations of alcohols before being cleared in xylene and washed in molten paraffin. These cell pellets were embedded in paraffin and arrayed in quadruplicate 1.0 mm cores using a manual tissue arrayer (MTA-1, Beecher Inc, WI). IHC was carried out on 5 µm sections.

#### **Immunohistochemical analysis**

Sections of cell pellet arrays or TMAs were deparaffinised in xylene and rehydrated in descending gradient alcohols. Heat-mediated antigen retrieval was performed using 10 mM sodium citrate buffer (pH 6.0) in a PT module (LabVision, UK) for 15 min at 95°C. The LabVision IHC kit (LabVision, UK) was used for staining. Endogenous peroxidase activity was blocked by incubation with 3% hydrogen peroxide for 10 min. Sections were blocked for 10 min in UV blocking agent. Rabbit polyclonal anti-human antibodies (HPA, Sweden) were applied at individual optimised dilutions for 1 hr (PDZK1 1:50

dilution; PBK, ANLN 1:150). Sections were washed in phosphate buffered saline with 0.1% Tween 20 (PBS-T). Subsequently, primary antibody enhancer was applied for 20 min, and sections were washed again in PBS-T. Sections were then incubated with HRP polymer for 15 min, washed in PBS-T and then developed for 10 min using diaminobenzidine (DAB) solution (LabVision, UK). After antigen retrieval, all incubations and washing stages were carried out at room temperature. The sections were counterstained in haematoxylin, dehydrated in alcohol and xylene and mounted using an automated coverslipper (Leica, Germany). As a negative control, the primary antibodies were substituted with PBS-T.

#### **Evaluation of immunohistochemical staining**

Slides were scanned at 20X magnification using a ScanScope XT slide scanner (Aperio Technologies, CA). Cores with less than 30% tissue present or less than 100 cells were discarded to avoid manual selection bias. Tumour samples were evaluated by at least two independent observers including one pathologist, and the maximum values of the two cores was used. All discordant cases were re-evaluated and a consensus reached between both observers. ANLN expression, as a nuclear marker, was categorised based on percentage nuclear staining such that 0 = ≤1%, 1 = 2-25%, 2 = 26-75% and 3 = > 75%. PDZK1 expression, as a cytoplasmic marker, was scored on a semi-quantitative scale depending on intensity of cytoplasmic staining: ranging from 0–3, where 0 is negative, 1 is weakly positive, 2 is medium positive and 3 is strongly positive. The intensity distribution (ID) scoring method was used with the cytoplasmic marker, PBK, which incorporated intensity of the scoring with percentage of cells stained [34].

#### **Annotation of gene expression data and hybridisation probes**

Gene expression data sets were downloaded from the Gene Expression Omnibus [35] or authors' websites in the form of raw data files where possible (Additional file 1: Table S2) [36-43]. Relevant gene expression and clinical data was extracted from ten publicly available datasets incorporating approximately 1,300 samples. Where raw data was not available, the normalised data as published by the original study was used. In the case of the Affymetrix datasets (.cel files), gene expression values were called using the robust multichip average method and data were quantile normalised using the Bioconductor package, affy [44,45]. For the dual-channel platforms, data were loess normalised using the Bioconductor package limma [46]. Hybridisation probes were mapped to Entrez gene IDs to gene-centre the data [47]. The Entrez gene IDs corresponding to the array probes targeting genes of interest were obtained from the Gene database at NCBI [48] (ANLN:54443, PBK:55872, PDZK1:5174). If there

were multiple probes for the same gene, the probes were averaged for that gene. All calculations were carried out in the R statistical environment [49].

#### **Statistical analysis of transcriptomic meta-analysis data**

Gene expression data from ten publicly available datasets were included in a meta-analysis to evaluate the individual prognostic significance of candidate proteins at the transcriptomic level, as previously described (Additional file 1: Table S2) [36-43]. Once a sample was assigned to a particular group, the 10 datasets were combined and a global survival analysis was performed. Each dataset was considered separately when determining which group a sample belonged to, due to the variability across different platforms. Recurrence-free survival (RFS) was considered the survival end point. Median mRNA levels established the cut-off for high and low expression for each biomarker. Survival curves of the dichotomised groups were compared using the log-rank test for significance. The survival curve was based on Kaplan-Meier estimates. Cox regression analysis was used to calculate hazard ratios (HR) and to adjust for all available clinical parameters. Across the meta-analysis, the available clinicopathological parameters were lymph node status, tumour grade and ER status.

#### **Statistical analysis of consecutive cohort data**

The  $\chi^2$  test and Fisher's exact test were used to evaluate associations between protein expression and clinicopathological variables in the cohort. Pearson's correlation coefficient was used to evaluate correlation between expression of the three independent markers. Kaplan-Meier analysis and the log-rank test were used to illustrate differences between recurrence-free survival (RFS) or breast cancer-specific survival (BCSS), according to differential protein expression. Cox proportional hazards regression was used to estimate proportional hazards for the individual protein expression and other clinicopathological variables in both univariate and multivariate models. The clinicopathological variables available for the consecutive cohort included tumour size, age at diagnosis, histological type, grade, nodal, ER, PR, Ki67 and Her2 status. All calculations were carried out using IBM SPSS Statistics version 20.0.

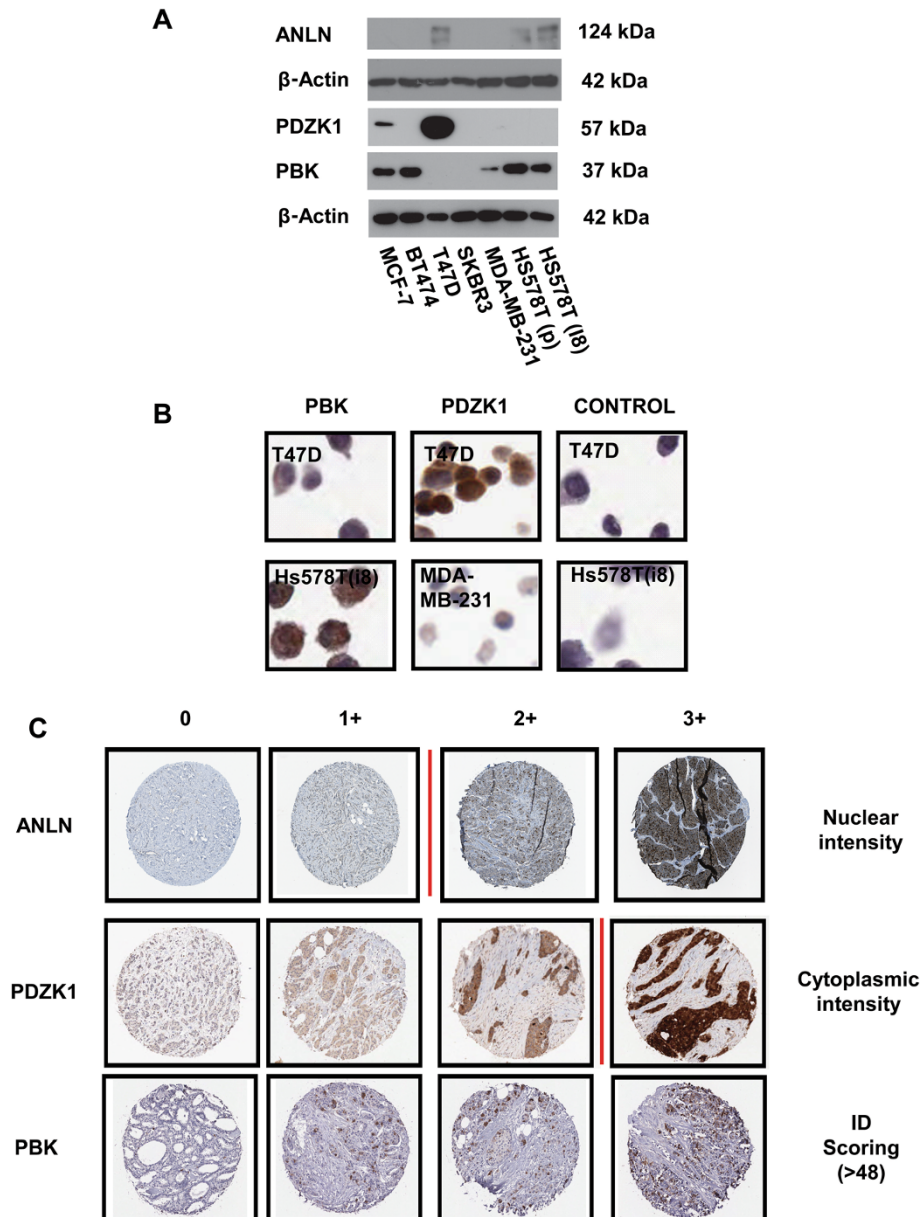
## **Results**

### **High-throughput screening platform for mono-specific antibodies against candidate breast cancer progression-related biomarkers**

In this study, fifty-six gene targets of interest were selected for generation of polyclonal affinity-purified anti-PrEST anti-sera on the basis of links with breast cancer progression at the mRNA level in previously published transcriptomic datasets [7,25,26]. Of the 56 gene targets submitted to the HPA, 18 mono-specific antibodies were

released for extended analysis. Specificity of the 18 antibodies was initially validated by Western blot analysis on a panel of discrete breast cancer cell lines with varying invasive properties. Ten out of the 18 antibodies exhibited specificity via Western blot analysis, with the expected molecular weight being observed (ANLN, PDZK1 and PBK shown in Figure 1A). Specificity was further verified by performing IHC on the corresponding

formalin-fixed, paraffin-embedded (FFPE) breast cancer cell lines (subset shown in Figure 1B). Seven antibodies showed concordant results for Western blot analysis and IHC staining in the breast cancer cell line cohort. Finally, three antibodies (PDZK1, ANLN, PBK) were successfully optimised on full-face paraffin embedded sections of breast cancer tissues and subsequently selected for screening on TMAs (Figure 1C).



**Figure 1 Expression of PBK, PDZK1 and ANLN protein in breast cancer.** **A:** Western blot analysis of PBK, PDZK1 and ANLN protein expression across a panel of 7 breast cancer cell lines of varying invasive capabilities. ANLN antibody specificity also validated by shRNA-mediated knockdown (data not shown). **B:** Validation of the PBK and PDZK1 antibodies by immunohistochemistry in a panel of FFPE breast cancer cell lines (x20 magnification). The T47D, MDA-MB-231 and Hs578T (i8) cell lines are specifically shown. Antibody positivity is indicated by the brown DAB staining. **C:** Representative cores of ANLN, PDZK1 and PBK protein expression from the TMAs graded on a scale from 0 to 3+ for protein staining intensity. Vertical red line represents the cut-off between low and high protein expression for each biomarker.

### Protein expression of lead candidate biomarkers in breast tumours determined by IHC

As shown in Figure 1A, antibodies against anillin (ANLN), PDZ-Domain Containing 1 (PDZK1) and PDZ-Binding Kinase (PBK) demonstrated specificity via Western blot analysis and exhibited concordant IHC staining on cell pellet arrays across 7 breast cancer cell lines. Antibody specificity was further validated using Western blotting and antigen microarrays (Additional file 2: Figure S1).

Four hundred and seventy-nine of the original cohort of 512 tumours (93.6%) were available for analysis of ANLN immunostaining, with 280/512 (54.7%) available for analysis of PDZK1 immunostaining and 292/512 (57.0%) available for analysis PBK immunostaining, with several sets of tumours not available for analysis due to core loss. Two hundred and fifty-two out of 512 (49.2%) tumours had a score for each of the three biomarkers, while 260 were not available for analysis due to core loss in the case of at least one of the 3 markers under evaluation. The clinicopathological variables for the available ( $n = 252$ ) and unavailable ( $n = 260$ ) tumours were compared using  $\chi^2$  analysis and Fisher's Exact test, with no significant difference being seen in terms of patient age ( $p = 0.927$ ), tumour size ( $p = 0.582$ ), tumour grade ( $p = 0.271$ ), histological type ( $p = 0.368$ ), nodal status ( $p = 0.479$ ), ER status ( $p = 0.578$ ), PR ( $p = 0.612$ ), Her2 ( $p = 0.192$ ) or Ki67 ( $p = 0.754$ ) expression between available and unavailable samples.

Using semi-quantitative analysis, IHC staining was scored on a scale of 0–3 based on intensity of staining (ANLN and PDZK1) or based on ID scoring (PBK) (see Figure 1C). High ANLN protein expression were classified as tumours with a staining intensity  $>1$ , and low expression classified as tumours with a staining intensity  $\leq 1$ . High PDZK1 protein expression was classified as tumours with a staining intensity  $>2$ , and low expression classified as tumours with a staining intensity  $\leq 2$ . PBK staining was classified using the ID scoring method (percentage of cells stained multiplied by intensity score), where the threshold for high PBK protein staining was  $>48$ . On the basis of this analysis, 309 evaluable tumours (64.5%) were classified as expressing high levels of ANLN and 170 (35.5%) expressing low levels of ANLN; 43 tumours (15.4%) were classified as expressing high levels of PDZK1 versus 237 (84.6%) expressing low levels of PDZK1, and 105 (36.0%) expressing high levels of PBK with 187 (64.0%) expressing low levels of PBK.

### Correlation of ANLN, PDZK1 and PBK protein expression with clinicopathological parameters

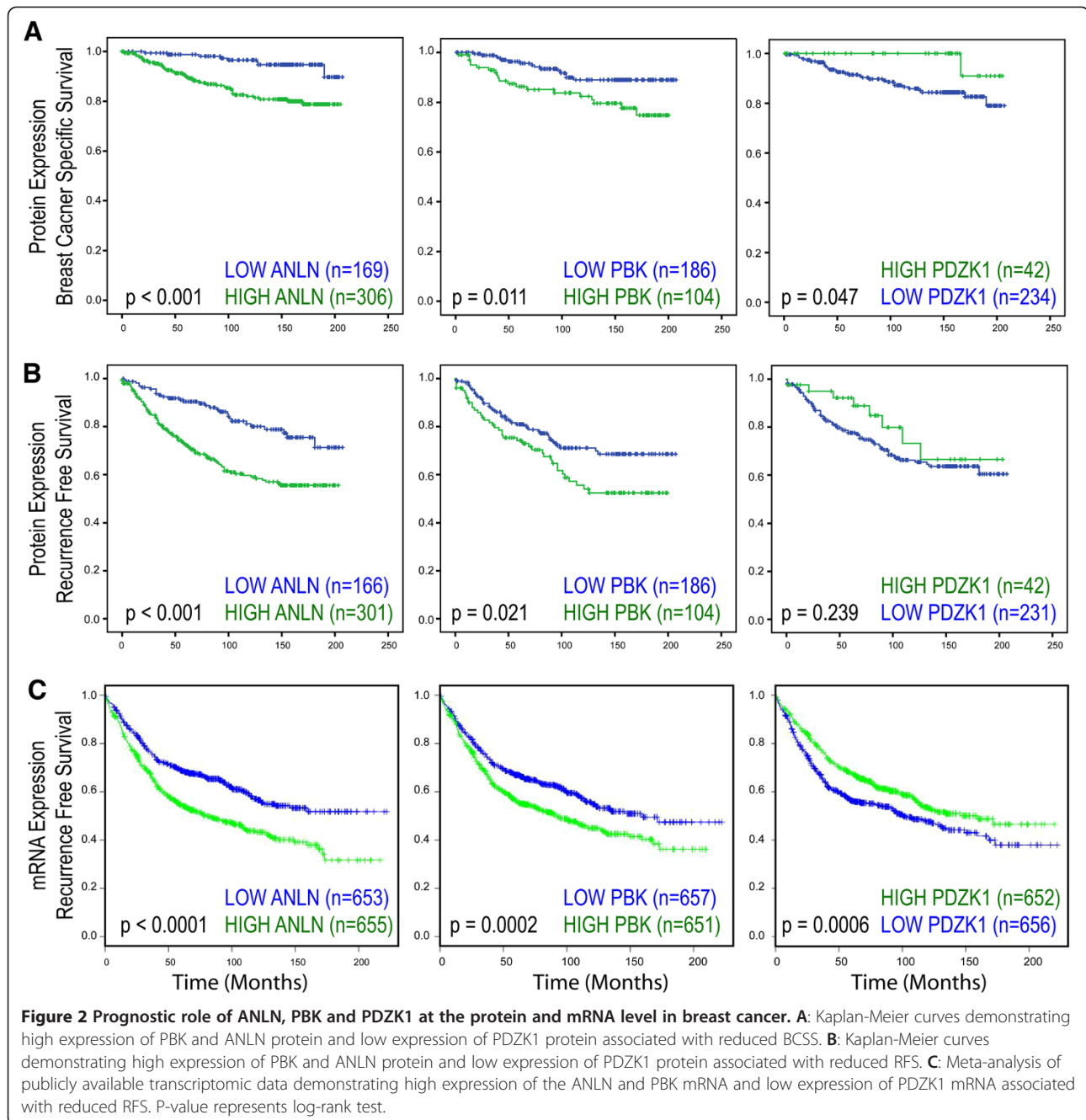
On the basis of the IHC thresholds for ANLN, PDZK1 and PBK expression detailed above, we investigated the associations between individual protein expression and a variety of well-defined clinicopathological variables in the TMA

cohort (Additional file 1: Table S3). ANLN expression correlated positively with tumour size ( $p = 0.006$ ), high tumour grade ( $p < 0.001$ ), Her2 status ( $p < 0.001$ ), Ki67 status ( $p < 0.001$ ) and invasive ductal carcinomas (IDC) ( $p < 0.001$ ), while correlating negatively with age at diagnosis ( $p = 0.019$ ), ER status ( $p < 0.001$ ) and PR status ( $p = 0.049$ ). PBK expression correlated positively with high grade tumours ( $p < 0.001$ ) and Ki67 status ( $p < 0.001$ ). PDZK1 expression correlated positively with low grade tumours ( $p = 0.010$ ). There was a significant correlation between ANLN and PBK expression (Pearson's  $R = 0.206$ ,  $p < 0.001$ ,  $n = 283$ ), yet there was no correlation between ANLN and PDZK1 ( $p = 0.410$ ), and PBK and PDZK1 ( $p = 0.543$ ).

### Single marker analysis of ANLN, PDZK1 and PBK protein expression associated with patient survival

The relationship between differential expression of ANLN, PDZK1 and PBK and outcome was subsequently examined. Kaplan-Meier analysis demonstrated that increased PDZK1 protein expression was associated with an improved BCSS ( $p = 0.047$ ), with high levels of ANLN and PBK protein expression being associated with reduced BCSS (ANLN:  $p < 0.001$ ; PBK:  $p = 0.011$ ) (Figure 2A). Univariate Cox regression analysis showed that high ANLN protein expression (HR = 3.91; 95% CI = 1.85–8.29;  $p < 0.001$ ) and high PBK protein expression (HR = 2.33; 95% CI = 1.19–4.55;  $p = 0.013$ ) were associated with reduced BCSS, while differential PDZK1 protein expression (HR = 0.17; 95% CI = 0.02–1.24;  $p = 0.080$ ) was not associated with prolonged BCSS. Both ANLN and PBK were significant independent predictors of BCSS when adjusted for other well-established variables, using multivariate Cox regression analysis (see Additional file 1: Table S4).

The relationship between ANLN, PBK and PDZK1 protein and RFS was examined. Kaplan-Meier analysis showed that high levels of ANLN and PBK protein expression being associated with reduced RFS (ANLN:  $p < 0.001$ ; PBK:  $p = 0.021$ ) (Figure 2B). PDZK1 protein expression was not associated with RFS ( $p = 0.239$ ). To compare the prognostic impact of ANLN with established factors, Cox regression analysis was performed. Univariate Cox regression analysis confirmed high ANLN expression (HR = 2.41; 95% CI = 1.61–3.62;  $p < 0.001$ ) and high PBK expression were associated with reduced RFS (HR = 1.64; 95% CI = 1.07–3.62;  $p = 0.023$ ). High PDZK1 expression was not associated with prolonged RFS (HR = 0.65; 95% CI = 0.31–1.35;  $p = 0.243$ ). In the multivariate Cox proportional hazards model, ANLN was a significant independent predictor of reduced RFS (HR = 2.14; 95% CI = 1.00–4.58;  $p = 0.038$ ). However, multivariate Cox regression analysis demonstrated that that PBK and PDZK1 protein expression were not independent predictors of RFS (Additional file 1: Table S5).

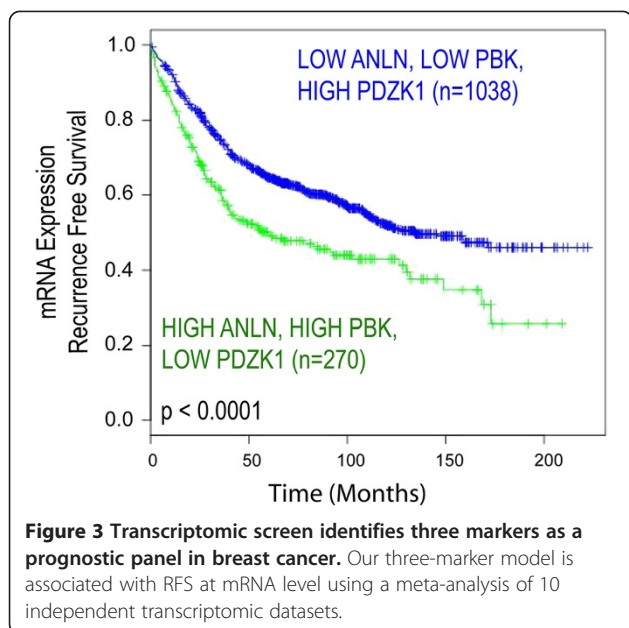


#### mRNA expression levels of ANLN, PDZK1 and PBK in a meta-analysis of publicly available breast cancer transcriptomic datasets

In order to validate these results in a larger number of patients, we performed a meta-analysis of ANLN, PDZK1 and PBK expression from independent transcriptomic datasets, previously described in detail (Additional file 1: Table S2) [36-43]. Using median mRNA expression levels as a cut-off, this meta-analysis displayed high concordance with protein expression data, whereby high expression of ANLN mRNA ( $p < 0.0001$ ), high expression of PBK

mRNA ( $p = 0.0002$ ) and low expression of PDZK1 mRNA ( $p = 0.0006$ ) were associated with decreased RFS (Figure 2C). This further confirms the role of ANLN and PBK as poor prognostic markers and PDZK1 as a good prognostic marker.

By combining these markers into a prognostic signature, we could test the strength of the panel depending on the relative expression of each marker. Patients with the poor prognostic signature (i.e. high expression of ANLN mRNA, high expression of PBK mRNA and low expression of PDZK1 mRNA) had reduced



RFS ( $p < 0.0001$ ,  $n = 1,308$ ) (Figure 3). Using Multivariate cox regression analysis and adjusting for known clinical parameters, these observations remained independent of lymph node status, tumour grade and ER status (HR = 1.49, 95% CI = 1.08-2.05,  $p = 0.018$ ,  $n = 699$ ).

**Correlation of 3-protein prognostic panel with clinicopathological parameters and patient survival**

Based on the above results, we combined ANLN, PDZK1 and PBK into a 3-protein prognostic model. A scoring technique was devised attributing a score of +1 to each poor prognostic marker exhibited on immunostaining of individual tumour cores (i.e., high ANLN, high PBK and low PDZK1) (Figure 1C). The summed scores for each patient ranged from 0–3 (good → poor prognosis) and, to establish a prognostic model, was dichotomised into a good prognostic set, ‘Signature A’ (incorporating scores 0 and 1), and a poor prognostic set, ‘Signature B’ (incorporating scores 2 and 3).

Two hundred and fifty-two patients of the original cohort had a score for all 3 candidate biomarkers. Associations of the established panel scores (0–3) to known clinicopathological parameters were assessed (Table 1). The panel score was found to correlate with high tumour grade ( $p < 0.001$ ), positive nodal status ( $p = 0.029$ ), ER-negativity ( $p = 0.006$ ), Her2-positivity ( $p = 0.036$ ) and high Ki67 status ( $p < 0.001$ ) status (Table 1).

When separated into 0, 1, 2 and 3 scores, a higher panel score was significantly associated with poorer BCSS and RFS (Figure 4A and 4C). The dichotomised 3 biomarker panel was significantly predictive of BCSS ( $p < 0.001$ ) (Figure 4B) and RFS ( $p < 0.001$ ) (Figure 4D). To compare the prognostic impact of the panel score with established

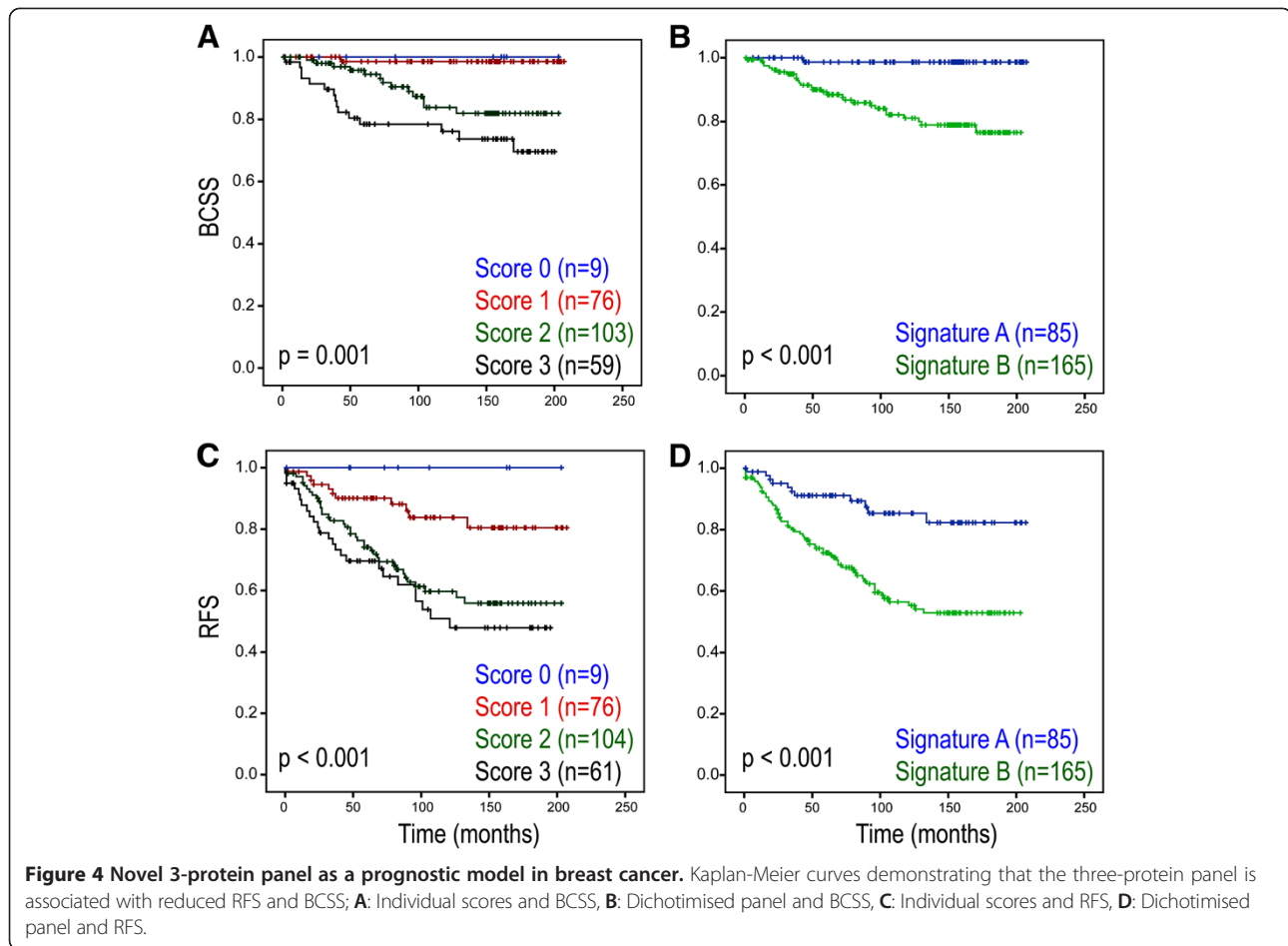
**Table 1 Association of panel score with clinicopathological parameters in the consecutive cohort**

Variables	Panel score				p-value
	0 (n = 9)	1 (n = 77)	2 (n = 105)	3 (n = 61)	
<b>Mean Age</b>					0.765
≤50	1 (11.1)	11 (14.3)	15 (14.3)	12 (19.7)	
>50	8 (88.9)	66 (85.7)	90 (85.7)	49 (80.3)	
<b>Tumour Size</b>					0.475
≤2cm	6 (66.7)	54 (70.1)	66 (62.9)	35 (57.4)	
>2cm	3 (33.3)	23 (29.9)	39 (37.1)	26 (42.6)	
<b>Histological type</b>					0.378*
Indeterminate	0 (0.0)	8 (10.4)	5 (4.8)	7 (11.5)	
Ductal	6 (66.7)	47 (61.0)	75 (71.4)	46 (75.4)	
Lobular	2 (22.2)	14 (18.2)	12 (11.4)	4 (6.6)	
Tubular	1 (11.1)	5 (6.5)	7 (6.7)	1 (1.6)	
Medullary	0 (0.0)	0 (0.0)	4 (3.8)	2 (3.3)	
Mucinous	0 (0.0)	3 (3.9)	2 (1.9)	1 (1.6)	
<b>Tumour Grade</b>					<0.001*
I	4 (44.4)	25 (32.9)	23 (21.9)	4 (6.6)	
II	5 (55.6)	42 (55.3)	38 (36.2)	19 (31.1)	
III	0 (0.0)	9 (11.8)	44 (41.9)	38 (62.3)	
<b>Nodal status</b>					0.029
N0	4 (66.6)	45 (68.2)	49 (51.0)	41 (73.2)	
N1+	2 (33.3)	21 (31.8)	47 (49.0)	15 (26.8)	
Unknown	3	11	9	5	
<b>ER status</b>					0.006
ER Negative	0 (0.0)	4 (5.3)	21 (20.6)	14 (23.7)	
ER Positive	8 (100)	72 (94.7)	81 (79.4)	45 (76.3)	
Unknown	1	1	3	2	
<b>PR status</b>					0.061
PR Negative	2 (28.6)	16 (26.2)	32 (37.2)	25 (51.0)	
PR Positive	5 (71.4)	45 (73.8)	54 (62.8)	24 (49.0)	
Unknown	2	16	19	12	
<b>Her2 status</b>					0.036
0 - 2 +	6 (85.7)	69 (97.2)	88 (87.1)	48 (81.4)	
3+	1 (14.3)	2 (2.8)	13 (12.9)	11 (18.6)	
Unknown	2	6	4	2	
<b>Ki67 status</b>					<0.001
0 - 10%	5 (62.5)	48 (62.3)	33 (32.7)	7 (11.7)	
11 - 100%	3 (37.5)	29 (37.7)	68 (67.3)	53 (88.3)	
Unknown	1	0	4	1	

\*Linear-by-linear  $\chi^2$  analysis; Others by Fisher's Exact test.

factors, Cox regression analysis was performed. Univariate Cox regression analysis demonstrated that high panel scores (2 and 3) were significantly associated with reduced BCSS (HR = 16.36; 95% CI = 2.23-120.30;  $p = 0.006$ ) and





reduced RFS (HR = 3.33; 95% CI = 1.75-6.31;  $p < 0.001$ ) (summarised in Table 2; all variables listed in Additional file 1: Table S6 and Additional file 1: Table S7). However, multivariate Cox regression demonstrated that the dichotomised 3-panel score was not a significant predictor of either BCSS (HR = 6.38; 95% CI = 0.79-51.26,  $p = 0.082$ ) or RFS (HR = 1.46; 95% CI = 0.66-3.19,  $p = 0.348$ ), when adjusted for other well-established variables, namely tumour grade, tumour size, age at diagnosis, ER, PR, Her2, Ki67 and nodal status. It must be noted when

all variables except for PR status are adjusted for, the 3-panel score becomes an independent predictor of BCSS (HR = 11.66; 95% CI = 0.150-90.68,  $p = 0.019$ ).

### Discussion

Gene expression profiling has successfully yielded new insights into the biologic diversity of breast cancer identifying several distinct molecular subtypes (such as luminal A, luminal B, basal and Her2) differing markedly in prognosis and in the repertoire of therapeutic targets they express

**Table 2 Cox univariate and multivariate analysis of RFS and BCSS in the consecutive cohort**

	BCSS		RFS	
	HR (95% CI)	<i>p</i>	HR (95% CI)	<i>p</i>
All patients (n = 252)				
3 marker panel	Univariate		Univariate	
Signature A	1.00		1.00	
Signature B	16.36 (2.23 - 120.30)	0.006	3.33 (1.75 - 6.31)	<0.001
3 marker panel	Multivariate*		Multivariate*	
Signature A	1.00		1.00	
Signature B	6.38 (0.79 - 51.26)	0.082	1.46 (0.66 - 3.19)	0.348

\* Multivariate analysis included adjustment for tumour size (continuous), tumour grade, age at diagnosis (continuous), nodal, ER, PR, Her2 and Ki67 status.

[4,5,50]. Importantly, these intrinsic subtypes play a key role in prediction of disease recurrence, treatment response, and the provision of new insights into oncogenic pathways and metastatic progression [51]. It is striking that, in the face of what is considered a heterogeneous tumour, molecular signatures of tumour subtypes consistently emerge across independent cohorts with diverse genetic and environmental backgrounds [52-54]. This reproducibility is a crucial primary descriptor of disease phenotype in the early detection of disease, lending key prognostic and predictive information.

Antibody-based proteomics occupies a pivotal space within the cancer biomarker discovery and validation pipeline, facilitating the high-throughput evaluation of candidate markers [17]. In this context, IHC-based high-throughput technology has been demonstrated as an effective platform for identification of protein surrogates of these intrinsic breast cancer subtypes by various groups [23,53]. For example, a panel of 5 proteins detected by immunohistochemistry was shown to be prognostic for ER-positive breast cancer [8]. The use of validated IHC surrogates should provide more clinically applicable assays in the future, due to ease of accessibility, low technical demand, cost-effectiveness and applicability to FFPE tissue. Despite these advances, the development of IHC-based assays has been globally impaired by the limited availability of high quality antibodies and lack of rigorous validation of emerging biomarkers. However, the development of comprehensive antibody resources and streamlining of reporting standards, promises to help overcome these obstacles [31,55].

In this study, we sought to determine whether insights from gene expression studies relating to breast cancer progression could be translated into a robust prognostic protein model using a discrete set of IHC markers. This proof-of-concept strategy generated a prognostic panel using high-throughput biomarker screening in combination with a devised panel scoring technique. We confirmed that a high panel score was significantly associated with reduced RFS ( $p < 0.0001$ ;  $n = 1,038$ ), using a meta-analysis of publicly available breast cancer transcriptomic datasets. The panel was an independent prognostic marker using multivariate Cox regression analysis ( $p = 0.018$ , HR = 1.49, 95% CI = 1.080-2.054,  $n = 699$ ). This strategy revealed a novel 3-marker prognostic model significantly predictive of RFS based on ANLN, PDZK1 and PBK expression patterns.

Next, we validated this signature on a protein-based platform using TMA technology. The 3-protein panel score correlated with known pathological prognostic variables, including tumour grade and lymph node status, ER, Her2 and Ki67 status. Univariate Cox regression analysis of RFS demonstrated that high panel scores, indicative of poor prognosis, were significantly associated with reduced RFS.

However, multivariate analysis demonstrated that the 3-marker panel score was not a significant predictor of either BCSS (HR = 6.38; 95% CI = 0.79-51.26,  $p = 0.082$ ) or RFS (HR = 1.46; 95% CI = 0.66-3.19,  $p = 0.348$ ), when adjusted for other well-established variables. We noted that the 3-panel score becomes an independent predictor of BCSS (HR = 11.66; 95% CI = 0.150-90.68,  $p = 0.019$ ), when all variables except for PR status are adjusted for. This may be due to marginal associations of our individual markers with these variables (e.g. PDZK1 and ER status:  $p = 0.041$ ; PDZK1 and PR status:  $p = 0.074$ ). Since both PDZK1 and PR are surrogate markers for ER activity, we note that the strength of this panel may be skewed by the presence of PDZK1 protein in the panel. Thus, we hypothesise that additional or alternate biochemical markers, interrogated using this high-throughput platform, may further augment the prognostic accuracy of this algorithm to a point that may allow implementation into routine clinical practice.

Interestingly, the 3 proteins that comprise this panel model are associated with distinct pathways in cancer biology. ANLN, initially characterised as a human homologue of anillin, a *Drosophila* actin-binding protein, is essential for the organisation of actin cables in the cleavage furrow, and plays a key role in cytokinesis and cell cycle progression [56-59]. ANLN has been demonstrated as a marker of poor prognosis, relating to aggressive cancer phenotypes [60]. In breast cancer, a transcriptomic study of DCIS to IDC breast cancer progression identified ANLN up-regulation in invasive tumour specimens relative to the pre-invasive phenotype [26]. Our study confirms the role of ANLN as a marker of poor prognosis, at the protein level, in an independent breast cancer cohort. PBK phosphorylates p38MAPK during mitosis, is considered a marker for cellular proliferation and is also implicated in DNA damaging sensing and repair [61,62]. PBK is associated with poorer prognosis in lung cancer [63], is up-regulated in IDC relative to DCIS at the transcriptomic level [26], and may be a promising molecular target for treatment of breast cancer [64]. Our findings further support the role of PBK as a marker of poor prognosis in breast cancer, with expression of PBK also associated with the histological markers of proliferation, Ki67 and tumour grade. PDZK1 is a known estrogen response gene in breast cancer, with proposed roles in signal transduction, cell polarity and ion exchange gating [65,66]. An in-house statistical re-analysis of the genes assessed by van't Veer and colleagues in the development the 70-gene prognostic signature identified PDZK1 as a marker of good prognosis in breast cancer [24], which we confirmed at the protein level in this study. The present study successfully validates these gene expression findings at the mRNA level, and also translates them at the protein level.

However, further studies are warranted at the *in vitro* and *in vivo* level, to help further interrogate the functional background of each of these markers in breast cancer progression. It will be necessary to further validate these findings with additional independent cohorts of samples to meet accepted international validation guidelines [55]. Although the literature is conflicting with regard to the best way to incorporate histopathology, IHC phenotypes, and gene expression data into an accurate classification system, our findings further support the key role of IHC prognostic models for current breast cancer management.

## Conclusions

We have developed a comprehensive biomarker development pathway, extending from discovery through to validation on TMAs, that can yield novel multi-protein panel signatures for use as a prognostic determinant in breast cancer. Such developments represent an important translational gateway into the era of individualised medicine for patients with newly diagnosed breast cancer.

## Additional files

**Additional file 1:** Supplementary tables.

**Additional file 2:** Supplementary figures.

## Abbreviations

ANLN: Anillin; BCA: Bicinchoninic acid; BCSS: Breast cancer-specific survival; BGA: Between group analysis; AUC: Area under the curve; DAB: Diaminobenzidine; DCIS: Ductal carcinoma *in situ*; ER: Estrogen receptor; FFPE: Formalin-fixed paraffin-embedded; Her2: Human epidermal growth factor receptor 2; His6ABP: Hexa-histidine albumin binding protein; HPA: Human protein atlas; HR: Hazard ratio; HRP: Horseradish peroxidase; ID: Intensity distribution; IDC: Infiltrating ductal carcinoma; IHC: Immunohistochemistry; p38MAPK: p38 mitogen-activated protein kinase; PBK: PDZ-binding kinase; PBS-T: Phosphate-buffered saline with 0.1% Tween 20; PDZK1: PDZ-Domain Containing 1; PR: Progesterone receptor; PrEST: Protein epitope signature tag; RIPA: Radioimmunoprecipitation assay buffer; ROC: Receiver operator curve; RFS: Recurrence-free survival; SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis; TBS-T: Tris-buffered saline with 0.1% Tween 20; TMA: Tissue microarray.

## Competing interests

The authors (Mathias Uhlén, Fredrik Pontén, Karin Jirstrom) currently hold a patent for the use of ANLN protein as an endocrine treatment predictive factor in breast cancer (Patent US20110269797).

## Authors' contributions

KJ coordinated the collection of patient tissue and constructed the tissue microarray. FP and MU managed the production of all antibodies. CK, SP and RTD carried out the Western blotting, cell pellet arrays and immunohistochemical analysis. POL, SM, DJB and ER performed the statistical analysis. POL and RTD helped to draft the manuscript. MJD, RZ, AHMc, MRK, KJ and WMG provided critical reading and revision of the manuscript. WMG and KJ conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors have read and approved the final manuscript.

## Acknowledgements

The authors wish to thank Dr. Peter Holloway and Dr. Osama A. Sharaf Eldin (UCD School of Medicine and Medical Science, University College Dublin,

Dublin 4, Ireland) for their invaluable assistance in the pathological scoring of the tumour tissue. Funding is acknowledged from the IRCSET Embark Initiative Postgraduate Scholarship Scheme (for support of Patrick O'Leary's work), the Eccles Breast Health Research Fund (for support of Dr. Roisin Dolan's post-graduate studies), Enterprise Ireland, Science Foundation Ireland (under the auspices of the 'Molecular Therapeutics for Cancer Ireland' Strategic Research Cluster) and the Health Research Board of Ireland (under the auspices of the 'Breast Cancer Metastasis: Biomarkers and Functional Mediators' research programme). The work of Dr. Zagodzón was funded by a Marie Curie International Reintegration Grant No. 224865 within the 7th European Community Framework Programme. Finally, the cross-national component of the project was facilitated by the Marie Curie Transfer of Knowledge Industry-Academia Partnership research programme, TargetBreast [www.targetbreast.com]. The UCD Conway Institute is funded by the Programme for Third Level Institutions (PRTL), as administered by the Higher Education Authority (HEA) of Ireland.

## Author details

<sup>1</sup>UCD School of Biomolecular and Biomedical Science, UCD Conway Institute of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland. <sup>2</sup>UCD School of Medicine and Medical Science, University College Dublin, Dublin 4, Ireland. <sup>3</sup>Molecular Therapeutics for Cancer Ireland, National Institute for Cellular Biotechnology, Dublin City University, Glasnevin, Dublin, Ireland. <sup>4</sup>Department of Pathology, University Hospital of Uppsala, Uppsala, Sweden. <sup>5</sup>Royal Institute of Technology, Stockholm, Sweden. <sup>6</sup>Department of Breast & Endocrine Surgery, Mater Misericordiae University Hospital, Dublin 7, Ireland. <sup>7</sup>UCD Clinical Research Centre, St. Vincent's University Hospital, Dublin 4, Ireland. <sup>8</sup>Division of Pathology, Department of Clinical Sciences, Lund University, Skåne University Hospital, Lund, Sweden.

Received: 21 November 2012 Accepted: 19 March 2013

Published: 2 April 2013

## References

1. Morris SR, Carey LA: **Molecular profiling in breast cancer.** *Rev Endocr Metab Disord* 2007, **8**(3):185-198.
2. Brennan DJ, Gallagher WM: **Prognostic ability of a panel of immunohistochemistry markers - retailoring of an 'old solution'.** *Breast Cancer Res* 2008, **10**(1):102.
3. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**(6797):747-752.
4. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**(19):10869-10874.
5. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Proc Natl Acad Sci USA* 2003, **100**(14):8418-8423.
6. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**(7403):346-352.
7. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**(6871):530-536.
8. Ring BZ, Seitz RS, Beck R, Shasteen WJ, Tarr SM, Cheang MC, Yoder BJ, Budd GT, Nielsen TO, Hicks DG, et al: **Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer.** *J Clin Oncol* 2006, **24**(19):3039-3047.
9. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, et al: **Converting a breast cancer microarray signature into a high-throughput diagnostic test.** *BMC Genomics* 2006, **7**:278.
10. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner F, Walker M, Watson D, Park T, et al: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**(27):2817-2826.
11. Sparano J: **TAILORx: trial assigning individualized options for treatment (Rx).** *Clin Breast Cancer* 2006, **7**(4):347-350.

12. Buysse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas A, d'Assignies M, Bergh J, Lidereau R, Ellis P, et al: **Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.** *J Natl Cancer Inst* 2006, **98**(17):1183–1192.
13. van't Veer LJ, Bernardis R: **Enabling personalized cancer medicine through analysis of gene-expression patterns.** *Nature* 2008, **452**(7187):564–570.
14. Brennan DJ, O'Brien SL, Fagan A, Culhane AC, Higgins DG, Duffy MJ, Gallagher WM: **Application of DNA microarray technology in determining breast cancer prognosis and therapeutic response.** *Expert Opin Biol Ther* 2005, **5**(8):1069–1083.
15. Larsson AH, Fridberg M, Gaber A, Nodin B, Levean P, Jonsson GB, Uhlen M, Birgisson H, Jirstrom K: **Validation of podocalyxin-like protein as a biomarker of poor prognosis in colorectal cancer.** *BMC Cancer* 2012, **12**(1):282.
16. Pascal LE, True LD, Campbell DS, Deutsch EW, Risk M, Coleman IM, Eichner LJ, Nelson PS, Liu AY: **Correlation of mRNA and protein levels: cell type-specific gene expression of cluster designation antigens in the prostate.** *BMC Genomics* 2008, **9**:246.
17. Brennan DJ, O'Connor DP, Rexhepaj E, Ponten F, Gallagher WM: **Antibody-based proteomics: fast-tracking molecular diagnostics in oncology.** *Nature Reviews Cancer* 2010, **10**(9):605–617.
18. Brennan DJ, Kelly C, Rexhepaj E, Dervan PA, Duffy MJ, Gallagher WM: **Contribution of DNA and tissue microarray technology to the identification and validation of biomarkers and personalised medicine in breast cancer.** *Cancer Genomics Proteomics* 2007, **4**(3):121–134.
19. Crabb SJ, Bajdik CD, Leung S, Speers CH, Kennecke H, Huntsman DG, Gelmon KA: **Clinically relevant prognostic subsets of breast cancer patients with four or more involved axillary lymph nodes be identified through immunohistochemical biomarkers? A tissue microarray feasibility study.** *Breast Cancer Res* 2008, **10**(1):R6.
20. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO: **Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype.** *Clin Cancer Res* 2008, **14**(5):1368–1376.
21. Caprioli RM: **Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy.** *Cancer Res* 2005, **65**(23):10642–10645.
22. Andersson AC, Stromberg S, Backvall H, Kampf C, Uhlen M, Wester K, Ponten F: **Analysis of protein expression in cell microarrays: a tool for antibody-based proteomics.** *J Histochem Cytochem* 2006, **54**(12):1413–1423.
23. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, et al: **Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma.** *Clin Cancer Res* 2004, **10**(16):5367–5374.
24. Culhane AC, Perriere G, Considine EC, Cotter TG, Higgins DG: **Between-group analysis of microarray data.** *Bioinformatics* 2002, **18**(12):1600–1608.
25. O'Brien SL, Fagan A, Fox EJ, Millikan RC, Culhane AC, Brennan DJ, McCann AH, Hegarty S, Moyna S, Duffy MJ, et al: **CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer.** *Int J Cancer* 2007, **120**(7):1434–1443.
26. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM, et al: **Gene expression profiles of human breast cancer progression.** *Proc Natl Acad Sci USA* 2003, **100**(10):5974–5979.
27. Svensson S, Jirstrom K, Ryden L, Roos G, Emdin S, Ostrowski MC, Landberg G: **ERK phosphorylation is linked to VEGFR2 expression and Ets-2 phosphorylation in breast cancer and is associated with tamoxifen treatment resistance and small tumours with good prognosis.** *Oncogene* 2005, **24**(27):4370–4379.
28. Lanigan F, Gremel G, Hughes R, Brennan DJ, Martin F, Jirstrom K, Gallagher WM: **Homeobox transcription factor muscle segment homeobox 2 (Msx2) correlates with good prognosis in breast cancer patients and induces apoptosis in vitro.** *Breast Cancer Res* 2010, **12**(4):R59.
29. Borgquist S, Jogi A, Ponten F, Ryden L, Brennan DJ, Jirstrom K: **Prognostic impact of tumour-specific HMG-CoA reductase expression in primary breast cancer.** *Breast Cancer Res* 2008, **10**(5):R79.
30. *Human Protein Atlas*. <http://www.proteinatlas.org>.
31. Uhlen M, Bjorling E, Agaton C, Szilyarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, et al: **A human protein atlas for normal and cancer tissues based on antibody proteomics.** *Mol Cell Proteomics* 2005, **4**(12):1920–1932.
32. Uhlen M, Ponten F: **Antibody-based proteomics for human tissue profiling.** *Mol Cell Proteomics* 2005, **4**(4):384–393.
33. Hughes L, Malone C, Chumsri S, Burger AM, McDonnell S: **Characterisation of breast cancer cell lines and establishment of a novel isogenic subclone to study migration, invasion and tumourigenicity.** *Clin Exp Metastasis* 2008, **25**(5):549–557.
34. Douglas-Jones AG, Schmid KW, Bier B, Horgan K, Lyons K, Dallimore ND, Moneypenny IJ, Jasani B: **Metallothionein expression in duct carcinoma in situ of the breast.** *Hum Pathol* 1995, **26**(2):217–222.
35. *Gene Expression Omnibus*. <http://www.ncbi.nlm.nih.gov/geo/>.
36. Calabro A, Beissbarth T, Kuner R, Stojanov M, Benner A, Asslaber M, Ploner F, Zatlouk K, Samonigg H, Poustka A, et al: **Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer.** *Breast Cancer Res Treat* 2009, **116**(1):69–77.
37. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, Dai H, He YD, van't Veer LJ, Bartelink H, et al: **Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival.** *Proc Natl Acad Sci USA* 2005, **102**(10):3738–3743.
38. Chanrion M, Negre V, Fontaine H, Salvetat N, Bibeau F, Mac Grogan G, Mauriac L, Katsaros D, Molina F, Theillet C, et al: **A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer.** *Clin Cancer Res* 2008, **14**(6):1744–1752.
39. Desmedt C, Giobbie-Hurder A, Neven P, Paridaens R, Christiaens MR, Smeets A, Lallemand F, Haibe-Kains B, Viale G, Gelber RD, et al: **The Gene expression Grade Index: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1–98 trial.** *BMC Med Genomics* 2009, **2**:40.
40. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, et al: **The molecular portraits of breast tumors are conserved across microarray platforms.** *BMC Genomics* 2006, **7**:96.
41. Kok M, Linn SC, Van Laar RK, Jansen MP, van den Berg TM, Delahaye LJ, Glas AM, Peterse JL, Hauptmann M, Foekens JA, et al: **Comparison of gene expression profiles predicting prognosis in breast cancer patients treated with tamoxifen.** *Breast Cancer Res Treat* 2009, **113**(2):275–283.
42. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, Gillet C, Ellis P, Ryder K, Reid JF, et al: **Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen.** *BMC Genomics* 2008, **9**:239.
43. Sabatier R, Finetti P, Cervera N, Lambaudie E, Esterni B, Mamessier E, Tallet A, Chabannon C, Extra JM, Jacquemier J, et al: **A gene expression signature identifies two prognostic subgroups of basal breast cancer.** *Breast Cancer Res Treat* 2011, **126**(2):407–420.
44. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**(2):249–264.
45. *Bioconductor package - affy*. <http://www.bioconductor.org>.
46. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30**(4):e15.
47. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2005, **33**:D54–58.
48. *Gene database at NCBI*. <http://www.ncbi.nlm.nih.gov/gene/>.
49. *The R Project for Statistical Computing*. <http://cran.r-project.org>.
50. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: **Breast cancer classification and prognosis based on gene expression profiles from a population-based study.** *Proc Natl Acad Sci USA* 2003, **100**(18):10393–10398.
51. Sotiriou C, Piccart MJ: **Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care?** *Nature reviews* 2007, **7**(7):545–553.
52. Anders CK, Deal AM, Miller CR, Khorram C, Meng H, Burrows E, Livasy C, Fritchie K, Ewend MG, Perou CM, et al: **The prognostic contribution of clinical breast cancer subtype, age, and race among patients with breast cancer brain metastases.** *Cancer* 2011, **117**(8):1602–11.
53. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, et al: **Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.** *JAMA* 2006, **295**(21):2492–2502.
54. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Gerads J, Bensen JT, et al: **Epidemiology of basal-like breast cancer.** *Breast Cancer Res Treat* 2008, **109**(1):123–139.
55. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG: **The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration.** *Ann Intern Med* 2003, **138**(1):W1–12.

56. Silverman-Gavrila RV, Hales KG, Wilde A: **Anillin-mediated targeting of peanut to pseudocleavage furrows is regulated by the GTPase Ran.** *Mol Biol Cell* 2008, **19**(9):3735–3744.
57. Piekny AJ, Glotzer M: **Anillin is a scaffold protein that links RhoA, actin, and myosin during cytokinesis.** *Curr Biol* 2008, **18**(1):30–36.
58. Hickson GR, O'Farrell PH: **Anillin: a pivotal organizer of the cytokinetic machinery.** *Biochem Soc Trans* 2008, **36**(Pt 3):439–441.
59. Hickson GR, O'Farrell PH: **Rho-dependent control of anillin behavior during cytokinesis.** *J Cell Biol* 2008, **180**(2):285–294.
60. Suzuki C, Daigo Y, Ishikawa N, Kato T, Hayama S, Ito T, Tsuchiya E, Nakamura Y: **ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway.** *Cancer Res* 2005, **65**(24):11314–11325.
61. Ayllon V, O'Connor R: **PBK/TOPK promotes tumour cell proliferation through p38 MAPK activity and regulation of the DNA damage response.** *Oncogene* 2007, **26**(24):3451–3461.
62. Hu F, Gartenhaus RB, Eichberg D, Liu Z, Fang HB, Rapoport AP: **PBK/TOPK interacts with the DBD domain of tumor suppressor p53 and modulates expression of transcriptional targets including p21.** *Oncogene* 2010, **29**(40):546–5474.
63. Shih MC, Chen JY, Wu YC, Jan YH, Yang BM, Lu PJ, Cheng HC, Huang MS, Yang CJ, Hsiao M, *et al*: **TOPK/PBK promotes cell migration via modulation of the PI3K/PTEN/AKT pathway and is associated with poor prognosis in lung cancer.** *Oncogene* 2012, **31**(19):2389–2400.
64. Fukukawa C, Ueda K, Nishidate T, Katagiri T, Nakamura Y: **Critical roles of LGN/GPSM2 phosphorylation by PBK/TOPK in cell division of breast cancer cells.** *Genes Chromosomes Cancer* 2010, **49**(10):861–872.
65. Kocher O, Krieger M: **Role of the adaptor protein PDZK1 in controlling the HDL receptor SR-BI.** *Curr Opin Lipidol* 2009, **20**(3):236–241.
66. Ghosh MG, Thompson DA, Weigel RJ: **PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer.** *Cancer Res* 2000, **60**(22):6367–6375.

doi:10.1186/1471-2407-13-175

**Cite this article as:** O'Leary *et al*: Systematic antibody generation and validation via tissue microarray technology leading to identification of a novel protein prognostic panel in breast cancer. *BMC Cancer* 2013 **13**:175.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

