J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 1

IJC
International Journal of Cancer

Additive effect of the AZGP1, PIP, S100A8 and UBE2C molecular biomarkers improves outcome prediction in breast carcinoma

Toshima Z. Parris[1], Anikó Kovács[2], Luaay Aziz[3], Shahin Hajizadeh[2], Szilárd Nemes[4,5], May Semaan[1], Eva Forssell-Aronsson[6], Per Karlsson[1] and Khalil Helou[1]

[1] Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Cancer Center, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[2] Laboratory of Clinical Pathology and Cytology, Sahlgrenska University Hospital, Gothenburg, Sweden
[3] Department of Otolaryngology, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
[4] Division of Clinical Cancer Epidemiology, Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden
[5] Regional Cancer Center (West), Western Sweden Health Care Region, Sahlgrenska University Hospital, Gothenburg, Sweden
[6] Radiation Physics, Institute of Clinical Sciences, Sahlgrenska Cancer Center, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

The deregulation of key cellular pathways is fundamental for the survival and expansion of neoplastic cells, which in turn can have a detrimental effect on patient outcome. To develop effective individualized cancer therapies, we need to have a better understanding of which cellular pathways are perturbed in a genetically defined subgroup of patients. Here, we validate the prognostic value of a 13-marker signature in independent gene expression microarray datasets (n = 1,141) and immunohistochemistry with full-faced FFPE samples (n = 71). The predictive performance of individual markers and panels containing multiple markers was assessed using Cox regression analysis. In the external gene expression dataset, six of the 13 genes (*AZGP1, NME5, S100A8, SCUBE2, STC2* and *UBE2C*) retained their prognostic potential and were significantly associated with disease-free survival ($p < 0.001$). Protein analyses refined the signature to a four-marker panel [AZGP1, Prolactin-inducible protein (PIP), S100A8 and UBE2C] significantly correlated with cycling, high grade tumors and lower disease-specific survival rates. AZGP1 and PIP were found in significantly lower levels in invasive breast tissue as compared with adjacent normal tissue, whereas elevated levels of S100A8 and UBE2C were observed. A predictive model containing the four-marker panel in conjunction with established clinical variables outperformed a model containing the clinical variables alone. Our findings suggest that deregulated AZGP1, PIP, S100A8 and UBE2C are critical for the aggressive breast cancer phenotype, which may be useful as novel therapeutic targets for drug development to complement established clinical variables.

Introduction

In 2008, breast cancer was the most commonly diagnosed malignancy and surpassed lung cancer as the leading cause of cancer-related death among females worldwide.[1] Therapeutic decisions for breast cancer are partially based on currently available predictive tests such as estrogen receptor (ER) status for endocrine response and HER2/*neu* status for HER2-targeted therapy. However, established prognostic factors [TNM staging (tumor size, extent of axillary lymph node involvement and metastasis), histological grade, menopausal status] still play an important role in the decision-making process by assessing the potential benefit of including adjuvant therapy in the treatment regimen. Current adjuvant systemic therapies are generally toxic and nonspecific, leading to potential over treatment of low-risk patients that may only receive modest benefit of treatment while under treating high-risk patients. To improve the management of breast cancer, we need to have a better understanding of the

Key words: breast cancer, outcome prediction, molecular biomarker, immunohistochemistry, model validation

Abbreviations: C-index: concordance index; CI: confidence interval; DFS: disease-free survival; DMFS: distant metastasis-free survival; DSS: disease-specific survival; ER: estrogen receptor; FFPE: formalin-fixed, paraffin-embedded tissues; GGI: genomic grade index; HR: hazard ratio; PgR: progesterone receptor; OS: overall survival

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Swedish Cancer Society, King Gustav V Jubilee Clinic Cancer Research Foundation, Percy Falk Research Foundation, Lion's Research Foundation, The Swedish Breast Cancer Association and the Wilhelm and Martina Lundgren Research Foundation
DOI: 10.1002/ijc.28497
History: Received 5 July 2013; Accepted 12 Sep 2013; Online 00 Month 2013

Correspondence to: Toshima Z. Parris, Department of Oncology, Sahlgrenska Cancer Center, University of Gothenburg, Medicinaregatan 1G, SE-41345 Gothenburg, Sweden, Tel.: 46-31-7866751, Fax: +46-31-820114, E-mail: toshima.parris@oncology.gu.se

ID: nareshrao Time: 14:10 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 2

What's new?

The development of new predictive tests to assess treatment and the identification of novel prognostic markers to predict outcome in patient subgroups could lead to significant improvements in breast cancer management. Here, the synergistic activity of four proteins, AZGP1, PIP, S100A8, and UBE2C, was found to serve as an effective marker for the stratification of breast cancer patients into risk groups for recurrence and death. The four-marker panel also improved outcome prediction when considered alongside established clinical variables. Overlapping signaling pathways between the proposed markers suggest that they may be attractive targets for breast cancer proteasome inhibitors.

predictive tests that can assess treatment response as well as prognostic markers that can predict outcome in a genetically defined subgroup of patients. In the last decade, genetic (DNA and RNA) and epigenetic (DNA methylation) profiling of breast cancer has revolutionized how we think by dramatically increasing our biological knowledge of the disease and opening new avenues for personalized treatment. However, to justify the routine clinical use of novel molecular markers, classifiers need to either provide additional information to established clinical variables or outperform them.

Previously, we identified a 13-gene prognostic signature for breast cancer by using genetic profiling (DNA and RNA).[2] These findings prompted us to test the hypothesis that deregulation of the 13 candidate biomarkers may perturb interconnecting cellular pathways, which have an adverse effect on clinical outcome. Although gene expression may give an indication of biological activity within cells, post-transcriptional regulation permits the translation of specific mRNA transcripts into protein products while other transcripts never reach the protein stage because these are pre-programmed for degradation. Therefore, it is essential that genetic analyses are performed at several biological levels (DNA–RNA–protein) to have a better understanding of which mechanisms cause gene deregulation in neoplastic cells (DNA copy number alterations, epigenetic modulation, etc.), whether aberrant gene expression levels are translated to aberrant protein expression levels, and ultimately how the phenotype is altered. In the current study, we validated our 13-marker signature using a test set comprised of publicly available gene expression datasets (n = 1,141). Then, we conducted a comprehensive study using immunohistochemistry on full-face formalin-fixed, paraffin-embedded tissues (FFPE) from the training set to provide further evidence that the 13 candidate biomarkers are biologically active in neoplastic cells. Lastly, we developed predictive models based on the relationship between clinical outcome and protein expression levels, using significant biomarkers alone and in combination with established clinical variables. Here, we present a combined predictive model for breast carcinoma containing a four-marker panel [AZGP1, Prolactin-inducible protein (PIP), S100A8 and UBE2C] together with established clinical variables that outperforms a model containing the clinical variables alone.

Material and Methods

Gene expression microarray datasets

To validate the 13-marker gene expression signature (*AZGP1, CBX2, DNALI1, LOC389033, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, SUSD3, STK32B* and *UBE2C*) in breast carcinoma, we used our previously published microarray data[2] as the training cohort (n = 97) and an independent cohort consisting of six publicly available Affymetrix U133A GeneChip datasets (n = 1,141) as the test cohort [Gene Expression Omnibus (GEO) accession numbers GSE1456, GSE2034, GSE4922, GSE6532, GSE7390 and GSE45255]. Data processing of the training and test sets was performed separately using Nexus Expression 3.0 (BioDiscovery).

Formalin-fixed paraffin-embedded tumor specimens

To correlate gene expression of the 13-marker signature with subsequent protein expression patterns in invasive breast tumor tissue, full-face formalin-fixed paraffin-embedded (FFPE) specimens were used. Multiple FFPE specimens corresponding to 71 of the 97 primary invasive breast cancer patients from the microarray training cohort were obtained from the Pathology Department at Sahlgrenska University Hospital in accordance with the Declaration of Helsinki and approved by the Medical Faculty Research Ethics Committee (Gothenburg, Sweden). Histological grading of the invasive tumor component was determined according to the Nottingham (Elston-Ellis) modification of the Scarff-Bloom-Richardson grading system.

Immunohistochemical analysis

Optimal antibody dilutions and assay conditions were achieved for immunohistochemistry using breast carcinoma as positive controls. Four micrometer FFPE sections were subsequently immunostained for AZGP1, CBX2, DNALI1, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, SUSD3, STK32B and UBE2C. Antibodies produced primarily by the human protein atlas (HPA) were chosen for use in the study, as HPA performs antibody specificity analyses using antigen microarrays. The sections were pretreated using the Dako PTLink system (Dako, Carpinteria, CA) and processed on an automated Dako Autostainer platform using the Dako Envision™ FLEX High pH Link Kit (pH 9) for rabbit anti-AZGP1 (Sigma-Aldrich HPA012582, 1:500 dilution), rabbit

ID: nareshrao Time: 14:10 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 3

anti-CBX2 (Sigma-Aldrich AV51628, 1:250 dilution), mouse anti-DNALI1 (Abcam ab58213, 1:25 dilution), rabbit anti-NME5 (ProteinTech Group 12923-1-AP, 1:400 dilution), rabbit anti-PIP (Sigma-Aldrich HPA009177, 1:25 dilution), rabbit anti-S100A8 (Sigma-Aldrich HPA024372, 1:2000 dilution), rabbit anti-SCUBE2 (Sigma-Aldrich HPA006353, 1:25 dilution), rabbit anti-SERPINA11 (Abcam ab86673, 1:50 dilution), rabbit anti-STC2 (ProteinTech Group 10314-1-AP, 1:500 dilution), rabbit anti-SUSD3 (Sigma-Aldrich HPA042310, 1:100 dilution), rabbit anti-STK32B (Sigma-Aldrich HPA015820, 1:50 dilution) and mouse anti-UBE2C (Abcam ab56861, 1:500 dilution). Peroxidase-catalyzed diaminobenzidine was used as the chromogen, followed by hematoxylin counterstain. The slides were then rinsed with deionized water, dehydrated in absolute alcohol, followed by 95% alcohol, cleared in xylene, and mounted. To facilitate histological assessment, one FFPE section was also stained with hematoxylin and eosin.

Immunostaining was evaluated by a breast pathologist, blinded to patient clinical outcome, and scored as previously described using the semiquantitative H-score method to calculate the sum of the percentage and intensity of positively stained invasive tumor cells (negative staining = 0; weak staining = 1+; moderate staining = 2+; strong staining = 3+). The H-score ranged from 0 to 300, where H-score = $(1 \times \%1+) + (2 \times \%2+) + (3 \times \%3+)$.[3] The X-tile software (version 3.6.1) was used to determine H-score cutoffs by dichotomizing patients according to H-score value and clinical outcome.[4] Scores were averaged over replicate FFPE sections representing the same tumor. Positive staining was interpreted as H-score >0 for S100A8, >10 for PIP and UBE2C and >50 for AZGP1. FFPE specimens lacking an invasive cancer component were excluded from the analysis.

Immunoblot analysis

Antibody specificity for the four-marker panel was tested further using immunoblot. Whole cell lysates were prepared using fresh-frozen tumor specimens in Mammalian Cell Lysis Buffer supplemented with protease inhibitor cocktail and Benzonase nuclease (Qiagen). Protein concentrations were determined using the Bradford Protein Assay (Bio-Rad). Protein extracts (50 μg) were resolved by SDS–PAGE on 4–12% bis-Tris gels (Invitrogen), transferred to nitrocellulose membranes (Fisher Scientific) probed with a primary anti-AZGP1 (Sigma-Aldrich, 1:500 dilution), anti-PIP (Sigma-Aldrich, 1:200 dilution), anti-S100A8 (Sigma-Aldrich, 1:1000 dilution) or anti-UBE2C antibody (Abcam, 1:200 dilution), followed by incubation with an appropriate horseradish-conjugated secondary antibody. The membranes were stripped and reprobed with an anti-ACTB antibody (Abcam Ab6276, 1:2000 dilution) as a loading control. The immunoblots were visualized using the SuperSignal West Femto Maximum Sensitivity Substrate kit (Pierce). Digitalized images were acquired using Fujifilm Luminescent Image Analyzer LAS-4000 and analysis performed with the Image Gauge v4.0 software.

Statistical analysis

Statistical analyses were performed using a 0.05 p-value cutoff in R/Bioconductor (version 2.15.0). All p-values are two-sided.

A. External validation of the 13-marker gene expression signature. Univariate Cox proportional hazard models were calculated for the 13 biomarkers in the gene expression signature using disease specific-survival (DSS) and disease-free survival (DFS) for the training and test cohorts, respectively. Statistically significant biomarkers ($p < 0.05$) were included in subsequent prognostic models, which were then used to stratify the training and test cohorts into risk groups (high- or low-risk) using average hierarchical clustering with Pearson correlation. Breast cancer survival rates were defined as a) the time from initial diagnosis to breast cancer-related death for DSS, b) time from initial diagnosis to first relapse (locoregional or distant) or breast cancer-related death for DFS, c) time from initial diagnosis to death from any cause for overall survival (OS); d) time from initial diagnosis to distant metastasis for distant metastasis-free survival (DMFS), and e) time from operative lesion removal to detection of tumor recurrence for recurrence-free survival (RFS). Survival rates were depicted with Kaplan–Meier curves and tested with log-rank test. Then, the relationship between clinico–pathological features and the risk groups was evaluated using two-tailed Fisher's exact test.

B. Development of a predictive model for disease-specific survival using protein expression. DSS survival rates at different protein expression levels for each protein were depicted with Kaplan–Meier curves and tested with log-rank test. Then, the relationship between clinico–pathological features and protein expression was evaluated using two-tailed Fisher's exact test, followed by the calculation of Spearman correlation coefficients (two-tailed) to establish the relationship between previously published microarray gene expression data[2] and IHC protein expression patterns. Multivariate analysis was conducted using the Cox proportional hazard model for DSS with stepwise selection to assess the predictive strength and additive accuracy of a four-marker panel containing combined AZGP1/PIP/S100A8/UBE2C protein expression after adjusting for established clinico–pathological features (age at diagnosis, number of positive axillary lymph nodes, histological grade, ER/PgR status, HER2/*neu* status, pathological tumor size). A concordance index (C-index) for the time-dependent area under the ROC curve [AUC(t)] was calculated to assess model predictive performance, varying from C-index = 0.5 (no predictive power) to C-index = 1 (perfect prediction). A risk score was calculated for each patient using the weighted linear combination of variables, dichotomized outcomes (*e.g.*, breast cancer survivors or nonsurvivors) and the combined expression of the four-marker panel. Patients missing expression values for one or more of the four markers was excluded in the analysis. A linear predictor (η) was calculated where high-risk patients had $\eta > 0$ and low-risk patients had $\eta < 0$.

ID: nareshrao Time: 14:10 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 4

Table 1. Univariate Cox proportional hazard regression models in the training and test cohorts

		Training cohort (n = 97)[1]		Test cohort (n = 1,141)[2]	
No.	Variables	Coefficient	p-value	Coefficient	p-value
1	AZGP1	−0.431	**<0.001**	−0.158	**0.001**
2	CBX2	0.783	**<0.001**	0.181	0.679
3	DNALI1	−0.488	**<0.001**	−0.174	0.081
4	LOC389033	−0.477	**<0.001**	ND	ND
5	NME5	0.783	**<0.001**	−0.330	**0.014**
6	PIP	−0.231	**<0.001**	−0.032	0.258
7	S100A8	0.233	**0.001**	0.073	**0.011**
8	SCUBE2	−0.313	**<0.001**	−0.091	**0.003**
9	SERPINA11	−0.572	**<0.001**	ND	ND
10	STC2	−0.321	**<0.001**	−0.135	**<0.001**
11	STK32B	−0.577	**<0.001**	−0.144	0.233
12	SUSD3	−0.478	**<0.001**	ND	ND
13	UBE2C	0.569	**<0.001**	0.344	**<0.001**

Statistically significant variables ($p < 0.05$) are displayed in bold text.
[1]Disease-specific survival (DSS) used for Cox regression model.
[2]Disease-free survival (DFS) used for Cox regression model.
Abbreviation: ND: Not determined.

Results

External validation of the 13-marker signature for breast carcinoma

To validate the 13-marker gene expression signature in an independent dataset, a test cohort was compiled from six publicly available datasets consisting of 1,141 breast carcinomas profiled using Affymetrix U133A GeneChips. Three (*LOC389033, SERPINA11* and *SUSD3*) of the 13 genes in the signature were not found on the Affymetrix platform and were therefore excluded in the analysis of the test cohort. The clinical relevance of the 13-marker signature was assessed using univariate Cox regression models, which revealed six markers (*AZGP1, NME5, S100A8, SCUBE2, STC2* and *UBE2C)* associated with DFS in the test cohort (Table 1). Average linkage hierarchical clustering with Pearson correlation stratified the training set (n = 97) into two groups using the six-marker signature. The first group, henceforth termed the high-risk group, contained 33 samples showing upregulation of *S100A8* and *UBE2C* and downregulation of *AZGP1, NME5, SCUBE2* and *STC2*. The second group, henceforth termed the low-risk group, contained 64 samples with predominantly inverse regulation of the six genes. Clustering of the test cohort also divided the samples into two main clusters (295 high-risk and 835 low-risk patients). However, 11 samples were classified outside of the two main risk groups, which could be attributed to disparate expression of the *S100A8* gene. Despite using two different microarray platforms, similar gene expression patterns were observed for the risk groups in both cohorts (Fig. 1).

Survival analysis showed that the six-marker signature was a predictor of DSS for the training cohort [Hazard ratio (HR), 3.12; 95% confidence interval (95% CI), 1.75–5.56; P = 5.15E-05], as well as DFS (HR, 2.27; 95% CI, 1.06–4.85; P = 0.032), OS (HR, 2.10; 95% CI, 1.20–3.67; P = 0.008), and DMFS (HR, 1.71; 95% CI, 1.17–2.51; P = 0.005) for the test cohort (Fig. 2). No significant relationship was found between RFS (HR, 1.13; 95% CI, 0.87–1.47; P = 0.37) and the six-marker signature. In addition, we observed that the high-risk group was associated with aggressive breast cancer features in both cohorts, *e.g.*, high histological grade, steroid hormone negativity (estrogen and progesterone-receptors), HER2/*neu*-positivity, triple negative status and the HER2/ER- and basal-like intrinsic subtypes (Table 2). Furthermore, the high-risk group was significantly associated with high S-phase fraction in the training cohort.

Aberrant AZGP1, PIP, S100A8 and UBE2C protein expression are predictive of disease-specific survival

To further validate the microarray results, protein expression for the 13-marker signature was examined in relation to DSS. Immunohistochemistry was performed for the 13-marker signature (*LOC389033* excluded) using full-face FFPE sections from 71/97 of the breast carcinoma specimens in the training microarray cohort. Six samples were excluded from further analysis because of lack of an invasive breast cancer component in the FFPE section. Our results show that AZGP1-negativity (HR, 0.273; 95% CI, 0.114–0.653; P = 0.002), PIP-negativity (HR, 0.313; 95% CI, 0.141–0.694; P = 0.003), S100A8-positivity (HR, 3.41; 95% CI, 1.47–7.88; P = 0.002), and UBE2C-positivity (HR, 3.22; 95% CI, 1.22–8.49; P = 0.012) in neoplastic breast tissue had an impact on patient outcome (Fig. 3). Hence, AZGP1- and PIP-positivity indicated a protective effect on patient outcome (HR < 1), whereas S100A8- and UBE2C-positivity (HR > 1) indicated an adverse effect. In the invasive component, cytoplasmic and membranous staining was observed for AZGP1 and PIP, whereas cytoplasmic and nuclear staining was observed for S100A8 and UBE2C. Using the four-marker panel, 28/65 patients were classified as high-risk patients and 31/65 patients as low-risk, whereas 6/65 were not characterized due to missing values for one or more of the four proteins. In agreement with the microarray results, we found that the patients most at risk for recurrence and therefore a more unfavorable clinical outcome had larger, high histological grade tumors with the Basal-like phenotype, high S-phase and PgR-negative status (Table 3 and Fig. 3). The aberrant protein expression patterns displayed by AZGP1, PIP, S100A8 and UBE2C were confirmed using Western blot (Supporting Information Fig. S1). No significant relationship was found between DSS and expression of CBX2, DNALI1, NME5, SCUBE2, SERPINA11, STC2, STK32B or SUSD3. A positive relationship was shown between the Illumina microarray and IHC results for AZGP1, PIP, S100A8 and UBE2C (r_S = 0.51), while a negative relationship was found for the

ID: nareshrao Time: 14:10 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 5

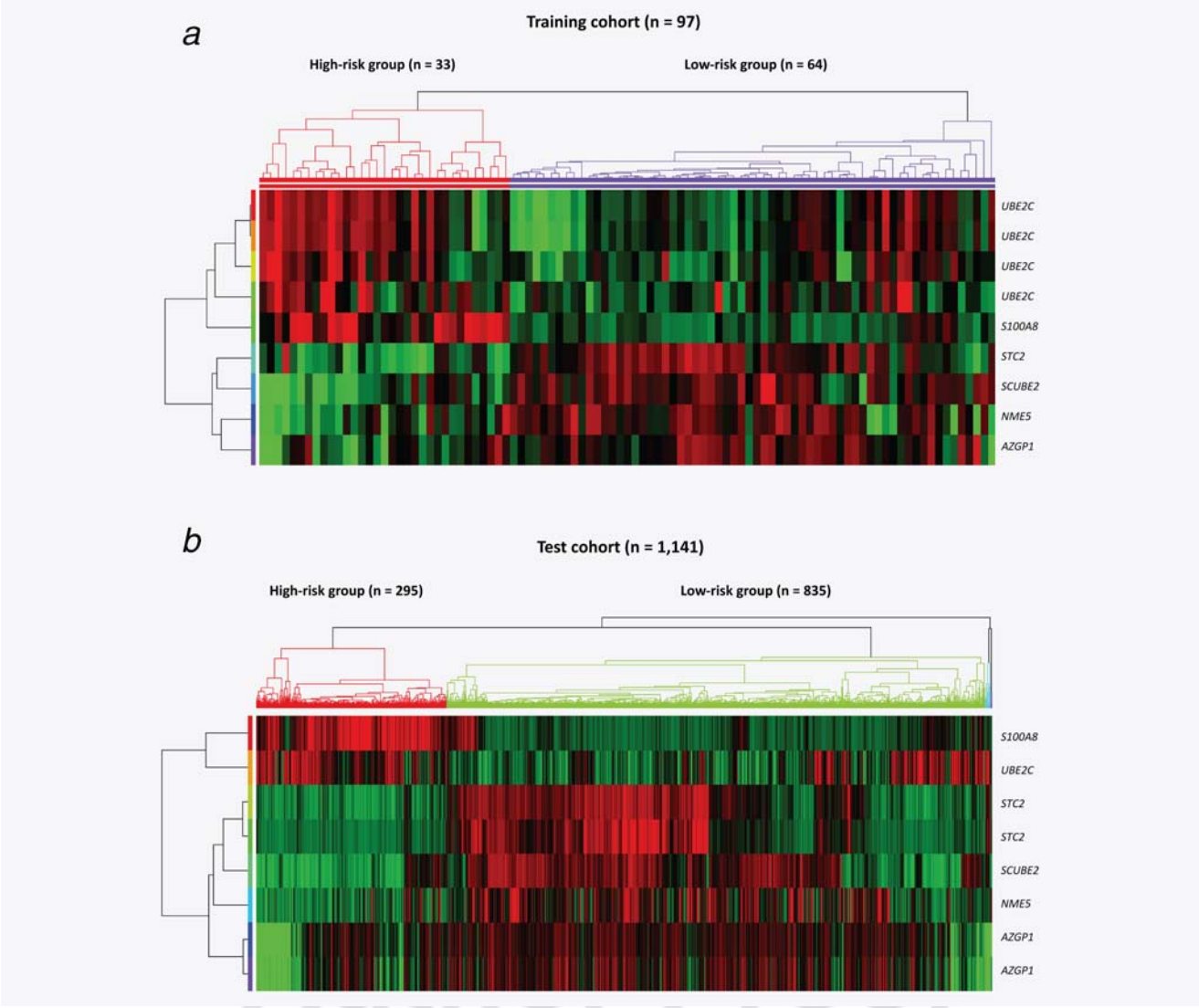


Figure 1. Two-dimensional cluster analysis of the six-marker signature in the training and test cohorts. (*a*, *b*) Classification of both cohorts into risk groups (high- and low-risk groups) in relation to gene expression of the six-marker signature. Each row represents a gene transcript and each column represents a tumor specimen. Upregulation is depicted as red color, downregulation as green and black as no change in gene regulation.

remaining candidate biomarkers ($r_S = -0.32$; Supporting Information Table S2).

A correlation analysis was performed between the four significant markers (AZGP1, PIP, S100A8 and UBE2C) and established clinico–pathological and molecular parameters [patient age at diagnosis, histological grade, Genomic Grade Index (GGI status), number of positive axillary lymph nodes, pathological tumor size, S-phase fraction, tumor inflammatory infiltration, steroid receptor status, HER2/*neu* status, triple negative status and breast cancer molecular subtype]. In general, a high fraction of cycling, high grade tumors were associated with AZGP1-negativity, PIP-negativity, S100A8-positivity and UBE2C-positivity. In addition, AZGP1 expression was significantly lower in triple-negative tumors with moderate tumor inflammatory infiltration and the basal-like subtype, whereas elevated levels of the S100A8 protein were also found in steroid receptor-negative tumors with strong tumor inflammatory infiltration (Table 3).

In agreement with the microarray results, we observed the frequent coexpression of AZGP1, PIP and UBE2C in the present dataset. Eighty-nine percent of AZGP1-negative tumors were PIP-negative ($p = 0.026$) and 70% of PIP-negative tumors were also UBE2C-positive ($p = 0.050$). Conversely, S100A8 was predominantly expressed independently of the other three proteins (AZGP1, PIP and UBE2C).

Outcome prediction is improved using the combined four-marker panel in conjunction with established prognostic markers

Having established the prognostic value of 4/12 analyzed candidate biomarkers (AZGP1, PIP, S100A8 and UBE2C) and their frequent coexpression, we evaluated whether a model

ID: nareshrao Time: 14:10 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 6

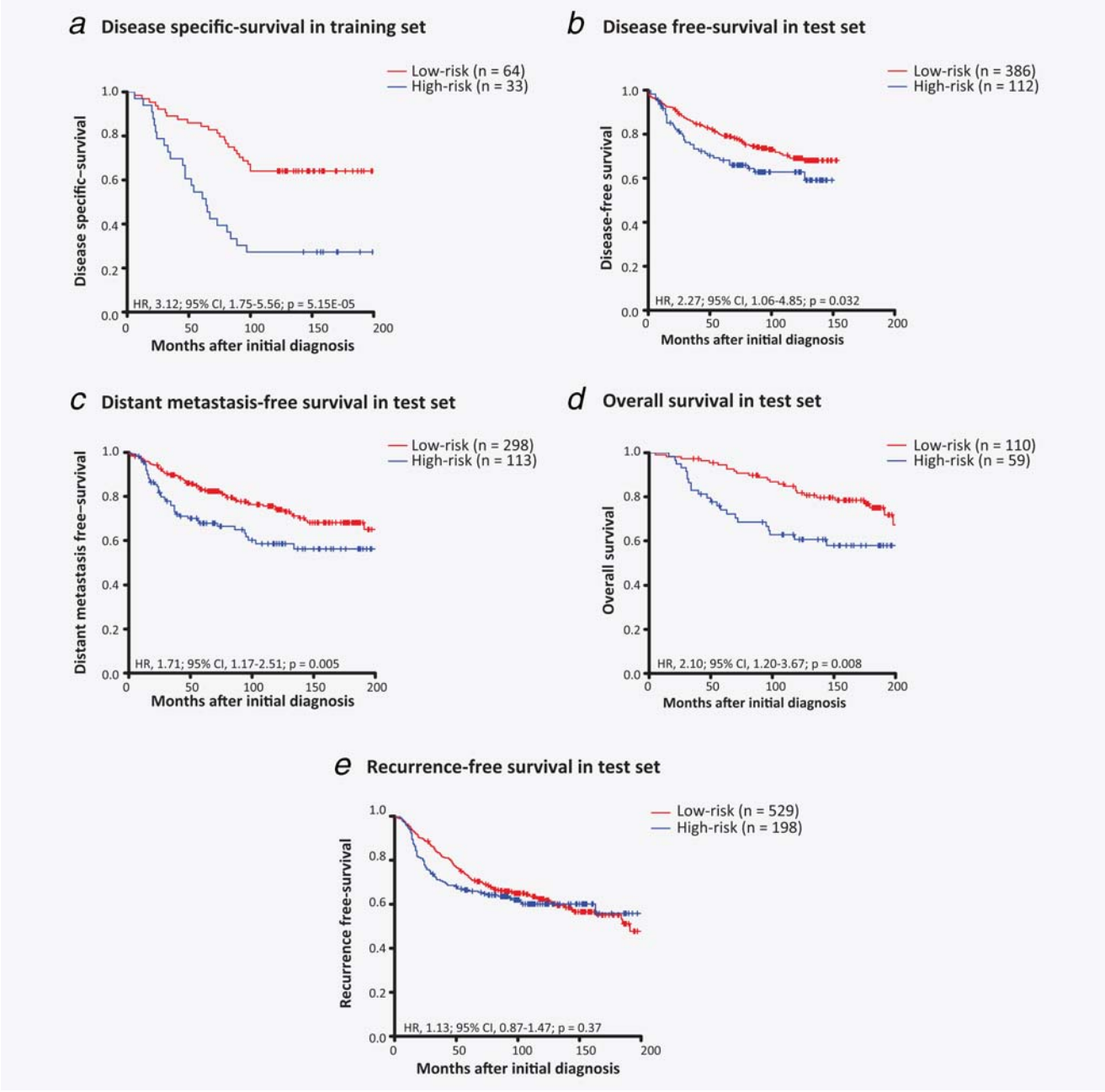


Figure 2. Kaplan–Meier analysis of the six-marker signature in the training and test cohorts. (*a*) Estimates of the probability of disease-specific survival according to risk assessment in the training cohort ($n = 97$). *p*-values, hazard ratios (HR), and 95% confidence intervals (95% CI) were calculated using the log-rank test and Cox proportional hazards regression, respectively. The *x*-axes depict months after initial diagnosis and the *y*-axes depict disease-specific survival. (*b–e*) Estimates of the probability of disease-free, distant metastasis-free, overall, and recurrence-free survival in the test cohort, respectively ($n = 1{,}141$). The *x*-axes depict months after initial diagnosis and the *y*-axes depict survival rates.

containing these four proteins could improve outcome prediction of breast carcinoma beyond established clinical variables (patient age at diagnosis, histological grade, number of positive axillary lymph nodes, pathological tumor size, ER/PgR status and HER2/*neu* status). First, we developed a model containing the four proteins, which proved to have more predictive power (C-index 0.735) than any of the four proteins alone (range, C-index 0.586-0.628). Furthermore, when a model was developed combining the four-marker panel together with established clinical variables, the C-index was increased from 0.773 (for the established clinical variables alone) to 0.836 (full model with all of the predictors). A stepwise multivariate analysis was then performed using the predictors from the full model. This analysis showed that a model containing AZGP1, PIP and S100A8 protein expression (3-marker panel) combined with the number of positive axillary lymph nodes, tumor size and PgR status as covariates ($p < 0.05$; C-index 0.826) performed similarly to the full model (Fig. 3).

ID: nareshrao Time: 14:10 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 7

Table 2. Clinicopathological features and risk assessment for breast carcinoma cases in the training and test cohorts

	No. of patients (%)					
	Training cohort ($n = 97$)			Test cohort ($n = 1,141$)		
Variables	High-risk patients ($n = 33$)	Low-risk patients ($n = 64$)	*p*-value	High-risk patients ($n = 295$)	Low-risk patients ($n = 835$)	*p*-value
Age			0.660			**0.010**
>50	20 (61)	41 (64)		92 (31)	339 (41)	
≤50	13 (39)	22 (34)		75 (25)	172 (21)	
Histological grade			**0.005**			**<0.001**
I	1 (3)	12 (19)		6 (2)	134 (16)	
II	16 (48)	32 (50)		68 (23)	336 (40)	
III	11 (33)	6 (9)		118 (40)	135 (16)	
GGI status			**<0.001**			ND
Low	3 (9)	36 (56)		-	-	
High	18 (55)	24 (38)		-	-	
No. of positive axillary lymph nodes			0.07			ND
0	16 (48)	31 (48)		-	-	
1–3	4 (12)	19 (30)		-	-	
≥4	13 (39)	14 (22)		-	-	
Pathologic tumor size			0.800			ND
pT1	7 (21)	19 (30)		-	-	
pT2	19 (58)	32 (50)		-	-	
pT3	5 (15)	10 (16)		-	-	
pT4	2 (6)	3 (5)		-	-	
S-phase fraction			**0.009**			ND
>6.1	13 (39)	9 (14)		-	-	
≤6.1	20 (61)	55 (86)		-	-	
Estrogen receptor status			**<0.001**			**<0.001**
Negative	12 (36)	4 (6)		150 (51)	73 (9)	
Positive	21 (64)	60 (94)		107 (36)	623 (75)	
Progesterone receptor status			**<0.001**			**<0.001**
Negative	24 (73)	13 (20)		29 (10)	29 (3)	
Positive	9 (27)	51 (80)		7 (2)	65 (8)	
HER2/neu status			**0.040**			**0.030**
Negative	26 (79)	60 (94)		13 (4)	55 (7)	
Positive	7 (21)	4 (6)		23 (8)	38 (5)	
Triple negative status			**<0.001**			**0.010**
Yes	10 (30)	1 (2)		9 (3)	7 (0.8)	
No	23 (70)	63 (98)		27 (9)	85 (10)	
Subtype			**<0.001**			**<0.001**
Luminal subtype A	0 (0)	1 (2)		1 (0.3)	38 (5)	
Luminal subtype B	18 (55)	60 (94)		2 (0.7)	21 (3)	
HER2/ER-	9 (27)	2 (3)		7 (2)	8 (0.1)	
Basal-like	6 (18)	1 (2)		18 (6)	7 (0.8)	
Normal-like	0 (0)	0 (0)		3 (1)	34 (4)	

p-values were calculated using the Fisher's exact test.
Abbreviation: ND, not determined.

ID: nareshrao Time: 14:11 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

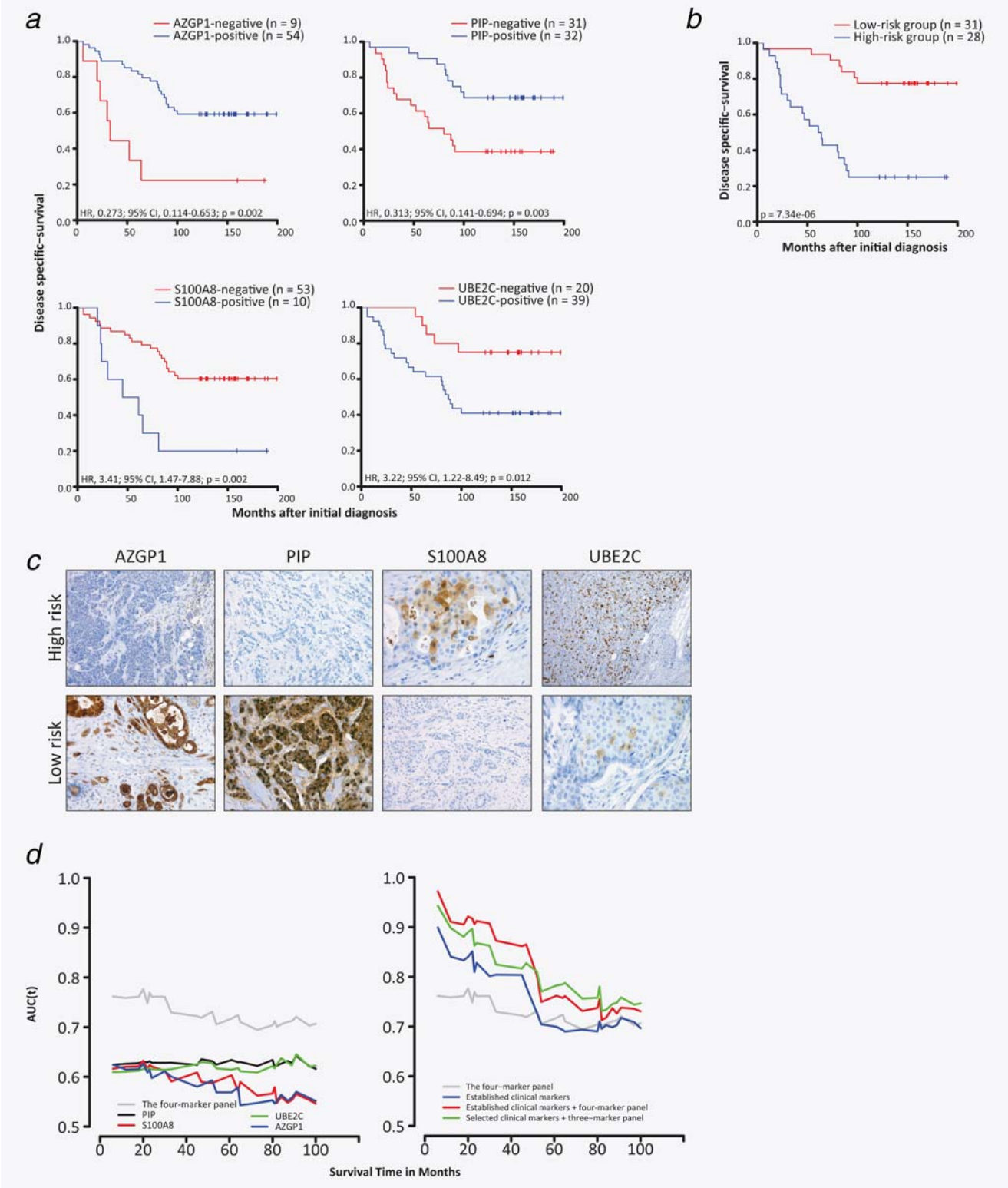


Figure 3. Prognostic potential of a four-marker panel in invasive breast carcinoma. (*a*) Kaplan–Meier estimates of the probability of disease-specific survival according to dichotomized protein expression for each biomarker in the panel (AZGP1, PIP, S100A8 and UBE2C). Patients with AZGP1-negative, PIP-negative, S100A8-positive and UBE2C-positive invasive breast tumors had significantly shorter survival times. *p*-values, hazard ratios (HR), and 95% confidence intervals (95% CI) were calculated using the log-rank test and Cox proportional hazards regression, respectively. The *x*-axes depict months after initial diagnosis and the *y*-axes depict disease-specific survival. (*b*) Patients stratified in the high-risk group had significantly shorter survival times than patients in the low-risk group. The *x*-axes depict months after initial diagnosis and the *y*-axes depict disease-specific survival. (*c*) Representative immunohistochemical staining showing protein expression in the invasive component of tumors stratified into low- and high-risk groups. (*d*) The lines represent the time-dependent area under the ROC curve [AUC(t)] for AZGP1, PIP, S100A8 and UBE2C protein expression assessed individually and combined as a four-marker panel. The estimated performance of the four-marker panel was better than that of either protein alone, increasing the C-index to 0.735; the predictive ability of the model including the four-marker panel was relatively stable over time. Combining established clinical variables (patient age at diagnosis, histological grade, number of positive axillary lymph nodes, pathological tumor size, ER/PgR status and HER2/*neu* status) with the four-marker panel further increased the C-index from 0.773 to 0.836. The *x*-axes depict survival time in months and the *y*-axes depict AUC(t).

Table 3. Relationship between AZGP1, PIP, S100A8 and UBE2C protein expression and clinicopathological features in breast carcinoma

		AZGP1 expression			PIP expression			S100A8 expression			UBE2C expression			AZGP1/PIP/S100A8/UBEC2 expression		
	Patients (n = 71)	AZGP1-positive (n = 54)	AZGP1-negative (n = 9)	p-value	PIP-positive (n = 32)	PIP-negative (n = 31)	p-value	S100A8-positive (n = 10)	S100A8-negative (n = 53)	p-value	UBE2C-positive (n = 39)	UBE2C-negative (n = 20)	p-value	Low-risk group (n = 31)	High-risk group (n = 28)	p-value
Age				0.140			0.190			1.00			0.170			0.430
>50	42 (59)	34 (63)	3 (33)		22 (69)	16 (52)		6 (60)	32 (60)		21 (54)	15 (75)		21 (68)	16 (57)	
≤50	28 (39)	19 (35)	6 (67)		9 (28)	15 (48)		4 (40)	20 (38)		17 (44)	5 (25)		10 (32)	12 (43)	
Histological grade				**0.038**			0.082			**0.001**			**0.020**			**0.010**
I	14 (20)	11 (20)	2 (22)		4 (13)	9 (29)		1 (10)	11 (21)		5 (13)	6 (30)		8 (26)	4 (14)	
II	35 (49)	36 (67)	3 (33)		23 (72)	15 (48)		3 (30)	35 (66)		23 (59)	14 (70)		22 (71)	15 (54)	
III	10 (14)	6 (11)	4 (44)		3 (9)	7 (23)		6 (60)	4 (8)		10 (26)	0 (0)		1 (3)	9 (32)	
GGI status				0.420			**0.050**			0.099			0.360			0.370
Low	27 (38)	22 (41)	2 (22)		17 (53)	8 (26)		1 (10)	24 (45)		12 (31)	9 (45)		14 (45)	7 (25)	
High	31 (44)	22 (41)	5 (56)		10 (31)	16 (52)		6 (60)	20 (38)		19 (61)	7 (35)		13 (42)	13 (46)	
No. of positive axillary lymph nodes				0.480			0.430			0.370			0.770			0.400
0	35 (49)	24 (44)	6 (67)		15 (47)	15 (48)		3 (30)	27 (51)		17 (44)	10 (50)		15 (48)	12 (43)	
1–3	18 (25)	16 (30)	1 (11)		10 (31)	7 (23)		3 (30)	14 (26)		10 (26)	6 (30)		10 (32)	6 (21)	
≥4	18 (25)	14 (26)	2 (22)		7 (22)	9 (29)		4 (40)	12 (23)		12 (31)	4 (20)		6 (19)	10 (36)	
Pathologic tumor size				0.760			0.097			0.640			0.250			**0.010**
pT1	16 (23)	12 (22)	2 (22)		10 (31)	4 (13)		1 (10)	13 (25)		5 (13)	7 (35)		11 (35)	1 (4)	
pT2	39 (55)	29 (54)	4 (44)		16 (50)	17 (55)		7 (70)	26 (49)		23 (59)	8 (40)		13 (42)	18 (64)	
pT3	12 (17)	10 (19)	2 (22)		6 (19)	6 (19)		2 (20)	10 (19)		8 (21)	4 (20)		6 (19)	6 (21)	
pT4	4 (6)	3 (6)	1 (11)		0 (0)	4 (13)		0 (0)	4 (8)		3 (8)	1 (5)		1 (3)	3 (11)	
S-phase fraction				**0.009**			**0.032**			0.420			**0.005**			**0.004**
>6.1	17 (24)	7 (13)	5 (56)		3 (9)	10 (32)		3 (30)	10 (19)		12 (31)	0 (0)		2 (6)	11 (39)	
≤6.1	54 (76)	47 (87)	4 (44)		29 (91)	21 (68)		7 (70)	43 (81)		27 (69)	20 (100)		29 (94)	17 (61)	
Tumor inflammatory infiltration				**0.005**			1.00			**0.0007**			1.00			0.110
Minimal	40 (56)	36 (67)	2 (22)		18 (56)	19 (61)		1 (10)	36 (68)		23 (59)	12 (60)		22 (71)	13 (46)	
Moderate	19 (27)	11 (20)	7 (78)		10 (31)	9 (29)		6 (60)	13 (25)		12 (31)	6 (30)		7 (23)	11 (39)	

Table 3. Relationship between AZGP1, PIP, S100A8 and UBE2C protein expression and clinicopathological features in breast carcinoma (Continued)

		AZGP1 expression			PIP expression			S100A8 expression			UBE2C expression			AZGP1/PIP/S100A8/UBEC2 expression		
	Patients (n = 71)	AZGP1-positive (n = 54)	AZGP1-negative (n = 9)	p-value	PIP-positive (n = 32)	PIP-negative (n = 31)	p-value	S100A8-positive (n = 10)	S100A8-negative (n = 53)	p-value	UBE2C-positive (n = 39)	UBE2C-negative (n = 20)	p-value	Low-risk group (n = 31)	High-risk group (n = 28)	p-value
Strong	6 (8)	6 (11)	0 (0)		3 (9)	3 (10)		3 (30)	3 (6)		3 (8)	2 (10)		1 (3)	4 (14)	
Estrogen receptor status				0.058			0.540			**0.017**			0.730			0.051
Negative	13 (18)	8 (15)	4 (44)		5 (16)	7 (23)		5 (50)	7 (13)		9 (23)	3 (15)		3 (10)	9 (32)	
Positive	58 (52)	46 (85)	5 (56)		27 (84)	24 (77)		5 (50)	46 (87)		30 (77)	17 (85)		28 (90)	19 (68)	
Progesterone receptor status				0.490			0.450			**0.001**			0.170			**0.020**
Negative	29 (41)	23 (43)	5 (56)		12 (38)	15 (48)		9 (90)	18 (34)		20 (51)	6 (30)		9 (29)	17 (61)	
Positive	42 (59)	31 (57)	4 (44)		20 (63)	16 (52)		1 (10)	35 (66)		19 (49)	14 (70)		22 (71)	11 (39)	
HER2/neu status				1.00			1.00			0.110			0.700			0.460
Negative	62 (87)	46 (85)	8 (89)		28 (88)	27 (87)		7 (70)	48 (91)		33 (85)	18 (90)		28 (90)	23 (82)	
Positive	9 (13)	8 (15)	1 (11)		4 (13)	4 (13)		3 (30)	5 (9)		6 (15)	2 (10)		3 (10)	5 (18)	
Triple negative status				**0.011**			0.470			0.110			0.700			0.130
Yes	9 (13)	4 (7)	4 (44)		3 (9)	5 (16)		3 (30)	5 (9)		6 (15)	2 (10)		2 (6)	6 (21)	
No	62 (87)	50 (93)	5 (56)		29 (91)	26 (84)		7 (70)	48 (91)		33 (85)	18 (90)		29 (94)	22 (79)	
Subtype				**0.003**			0.250			0.160			0.240			**0.010**
Luminal **subtype B**	57 (80)	46 (85)	4 (44)		27 (84)	23 (74)		6 (60)	44 (83)		29 (74)	18 (90)		29 (94)	18 (64)	
HER/ER-	8 (11)	6 (11)	1 (11)		4 (13)	3 (10)		2 (20)	5 (9)		5 (13)	2 (10)		2 (6)	5 (18)	
Basal-like	6 (8)	2 (4)	4 (44)		1 (3)	5 (16)		2 (20)	4 (8)		5 (13)	0 (0)		0 (0)	5 (18)	

p-values were calculated using the Fisher's exact test.

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 11

Discussion

In a previous retrospective study, we identified a 13-marker prognostic signature in invasive breast tumors with aggressive features.[2] Here, we evaluated the performance of this signature in independent gene expression microarray datasets from publicly available breast cancer cohorts (n = 1,141) and assessed the applicability of the signature as immunohistochemistry markers in relation to clinically relevant breast cancer features using full-face FFPE samples from the training set (n = 71). Despite difficulties evaluating the full 13-marker signature in the test cohort due to differences in the microarray platforms (Illumina *vs.* Affymetrix), similar expression patterns were found for the evaluated genes in both the training and test cohorts. In addition, Cox regression models demonstrated the continued prognostic potential of 6/10 evaluated genes in the external cohort. The prognostic ability for *CBX2*, *DNALI1*, *PIP* and *STK32B* may have differed in the test set for several reasons, *e.g.*, the difference in target probe sequences on the different microarray platforms, the relatively small sample size for the training set, and the use of different endpoints (DSS for the training set and DFS for the test set).

Interestingly, we observed a partial overlap between the 13-marker signature described here and other breast cancer risk assessment signatures, *i.e.*, MammaPrint®, Oncotype Dx™, PAM50, EndoPredict, Genomic Grade Index and the Invasiveness gene signature (Supporting Information Table S1); the *S100A8* gene was the only marker in the six-marker signature from the gene expression validation study (*AZGP1, NME5, SCUBE2, STC2, S100A8* and *UBE2C)* that has not been previously identified in other outcome predictors.[5–12] There is, however, an important disadvantage in using gene expression microarrays; although mRNA expression may provide an indication of biological activity, mRNA and protein expression levels do not always correlate. Evaluation of the biomarker signature in external gene expression datasets and then using immunohistochemistry in conjunction with clinico–pathological variables has provided further evidence that at least four of the proposed markers (AZGP1, PIP, S100A8 and UBE2C) are clinically relevant and biologically active in neoplastic cells. In the final four-marker signature (AZGP1, PIP, S100A8 and UBE2C), three markers were also present in the six-marker gene expression signature (*AZGP1*, *S100A8* and *UBE2C)*. PIP activity is induced transcriptionally by androgens and post-transcriptionally by prolactin, which may explain why PIP was identified as a significant marker in the predictive marker signature using protein expression but not in the six-marker gene expression signature.[13] Taken together, we demonstrated the prognostic value of aberrant protein expression levels for the four-marker panel (AZGP1, PIP, S100A8 and UBE2C) in invasive breast cancer and showed that this phenotype is associated with cycling, high histological grade tumors. Furthermore, a predictive model containing the four-marker panel coupled with established clinical variables surpassed the performance of a model using the established clinical variables alone. However, the four-marker panel will also need to be validated using immunohistochemistry with an independent breast cancer cohort to further establish its clinical utility.

Few gene expression signatures for prognosis of breast cancer have also been evaluated at the protein level. Gene expression microarrays are a relatively quick and inexpensive method that reveals a snapshot of which genes are expressed in cells or a tissue mass at a specific time point. However, if the different cell types within a heterogeneous tissue such as breast carcinoma (*e.g.*, malignant, nonmalignant, normal, stroma, etc.) are not analyzed separately, expression levels within the tumor mass will either be over- or underestimated. Here, a comparison between the protein expression patterns in neoplastic cells and the gene expression profiling analyses (the collective expression of all cell types within the tumor mass) yielded a higher correlation coefficient for AZGP1, PIP, S100A8 and UBE2C protein expression than for the remaining candidate biomarkers. The moderate correlation between the two studies may be partially due to the differences in expression levels for specific genes and corresponding proteins, the importance of analyzing different cell types separately, as well as, the choice of antibody, which may have low specificity for a particular antigen or may not be specific for the isoform of interest. The majority of the antibodies (AZGP1, PIP, S100A8, SCUBE2, SUSD3 and STK32B) used in the present study were chosen from the HPA, which performs antibody specificity analyses using antigen microarrays. SUSD3 was the only antibody with low specificity among the antibodies chosen from HPA. To prevent signal saturation or weak bands, the optimal antibody dilutions (for PIP, S100A8 and UBE2C) used for immunoblots differed slightly from that used for immunohistochemistry. Differences in sample preparation (*e.g.*, fixation, pH, temperature, solvent composition and incubation time) may have influenced the availability of the antigen in both techniques and therefore the need for varying antibody dilutions. For immunoblot, the observed tissue protein expression levels may have resulted from a contribution of both neoplastic cells and surrounding normal tissue.

Zinc-alpha2-glycoprotein (AZGP1) is a secretory protein found in many body fluids,[14–17] normal exocrine glandular epithelia of various normal tissues,[16] adipose tissue,[18] benign diseases[19] and cancer.[20–25] In normal tissues, AZGP1 plays a role in increased susceptibility to obesity[26] and adipocyte differentiation when induced by PPARγ nuclear receptor and inhibited by TNFα.[27] In cancer, elevated levels of AZGP1 were found in the urine of patients with cancer cachexia and the protein was therefore proposed to be involved in cancer-associated lipolysis.[17] A study conducted by He *et al.* showed a reduction in cellular proliferation and inhibition of cdc2, a cyclin-dependent kinase that regulates the G2/M transition, in AZGP1-stimulated tumor cells.[28] In pancreatic cancer, it was shown that lower levels of AZGP1 were the result of histone acetylation followed by aggressive tumor features and induction of epithelial-mesenchymal transition, whereas the epithelial phenotype was maintained in epigenetically

ID: nareshrao Time: 14:12 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

J_ID: z3q Customer A_ID: IJC28497 Cadmus Art: IJC28497 Ed. Ref. No.: 13-1726.R2 Date: 3-October-13 Stage: Page: 12

regulated AZGP1-expressing cells.[29] This finding may explain why a high fraction of basal-like tumors lacked expression of the AZGP1 protein (67%) in this study.

Prolactin-inducible protein (PIP) is a secreted monomer glycoprotein that is expressed in exocrine glands, gross cystic disease and breast cancers exhibiting apocrine features.[30] T47D, where the *PIP* gene is duplicated as a palindrome of the 7q34–q35 genomic region, and recently VHB1 breast cell lines are among the only cells, which show the suppression of cell proliferation and migration as well as the induction of apoptosis in androgen-stimulated PIP-expressing cells.[31] Androgens also stimulate the release of AZGP1 and apoD in T47D cells. In addition to androgens, PIP expression is also regulated by other steroids and lactogenic hormones produced by the pituitary gland. In a study by Haagensen *et al.*, progesterone stimulation resulted in the release of PIP and AZGP1, but had no significant effect on apoD levels, whereas estradiol had no effect on any of the three proteins.[32] Consequently, these three proteins are also major components of human breast secretions, such as gross cystic fluid, and seminal plasma.[19,33] Recently, PIP induction was found to be initiated by the synergistic activity of the androgen receptor and RUNX2, a prometastatic transcription factor.[34] In breast cancer metastases of the bone, Jones *et al.* showed that RUNX2 expression could be suppressed by treatment with 26S proteasome inhibitor bortezomib.[35] Nevertheless, PIP expression has been observed in a high fraction of breast tumors (12–85%) and associated with ER-positivity, PgR-positivity, low tumor grade and relapse-free survival, whereas expression of the protein was frequently lost in advanced-stage tumors.[36–38]

S100 calcium-binding protein A8 (S100A8, calgranulin A) is a member of the S100 superfamily containing small, acidic proteins (~10 kDa) with the calcium-binding EF-hand motif, which can form a homodimer with itself and a 36 kDa heterocomplex (referred to as calprotectin) with S100A9 (calgranulin B) in a Ca^{2+}-dependent manner. Several of the S100 genes form a gene cluster on the 1q21 genomic region which are frequently deleted, translocated or duplicated in cancer. Both S100A8 and S100A9 function as proinflammatory cytokines and are expressed in cells of myelomonocytic lineage (*e.g.*, granulocytes, monocytes), macrophages, neutrophils, keratinocytes and advanced stage cancer and surrounding stroma, which may explain why 90% of S100A8-positivity was found in the presence of inflammation in the present study.[39–41] Interestingly, S100A8/A9, PIP and AZGP1 were all found in high amounts in human saliva.[42] In breast cancer, elevated levels of S100A8/A9 have been found in high grade tumors, estrogen-receptor negative tumors and tumors with the Basal-like phenotype.[43] S100A8 has also been implicated in the stimulation of HIV production in cervico-vaginal secretions.[44] Recently, Moon *et al.* showed that both S100A8 and S100A9 are involved in H-Ras-mediated cell invasion and migration.[45]

Cell cycle progression requires activation of the anaphase-promoting complex (APC), which triggers the metaphase-to-anaphase transition and mitotic exit by promoting ubiquitin-dependent proteolysis of mitotic cyclins. Ubiquitination of cyclins is a three-step process where E1 enzymes activate ubiquitin, E2 enzymes such as ubiquitin-conjugating enzyme E2C (UBE2C) conjugate and transfer ubiquitin molecules to the E3 enzyme, which transfers the ubiquitin to the target protein. Destruction of ubiquitinated cyclins is then carried out by proteasomes. Townsley *et al.* showed that ubiquitination of cyclins is followed by cdc2 inactivation. *In vitro*, dominant-negative UBE2C and suppression of UBE2C were shown to inhibit ubiquitination of cyclin A and B and thereby resulted in the accumulation of cells in mitosis and inhibition of anaphase onset.[46] The elevated levels of UBE2C found in several cancer forms have been associated with HER2-positivity, intense Ki-67 staining, and unfavorable clinical outcome.[46–48] In colorectal carcinoma, inhibition of UBE2C could be achieved using bortezomib, which in turn suppressed cell proliferation and disruption of the cell cycle. UBE2C overexpression observed in primary colon tumors and liver metastases as well as anaplastic thyroid carcinoma cell lines was attributed to gene amplification.[49,50]

In summary, we propose a prognostic model containing a four-marker panel (AZGP1, PIP, S100A8 and UBE2C) in combination with established clinical variables. Although extensive research has been performed on AZGP1, PIP, S100A8 and UBE2C as individual proteins, this is the first report showing the addictive effect of the four proteins together in any cancer form. The mechanism(s) by which these four proteins play a role in breast cancer is not yet known. We have previously shown that gene deregulation of *AZGP1, PIP, S100A8* and *UBE2C* is not due to abnormal DNA copy number; S100A8 may be activated by several different mechanisms in cancer as 1/10 and 2/10 S100A8-positive samples showed DNA amplification (log2ratio ≥ 0.5) or gain (log2ratio $0.2 \geq x < 0.5$) of the *S100A8* gene, whereas none of the other genes showed DNA deletion or amplification.[2] However, several of these proteins share common functions, *e.g.*, AZGP1 and PIP are both secreted glycoproteins found in exocrine glands; AZGP1 and UBE2C are both involved in cell cycle regulation and thus have an effect on cell proliferation and inhibition of cdc2 activity; AZGP1, PIP and S100A8 are all found at high amounts in human fluids which contain proteins involved in inflammatory and immune responses; FDA approved proteasome inhibitor bortezomib targets UBE2C and PIP expression. In addition, protein–protein interaction studies show that these four proteins belong to overlapping signaling pathways. Taken together, further studies are warranted to establish the role these four markers have in cancer progression and whether the four-marker panel may be useful in the decision-making process of breast cancer using an independent cohort.

Acknowledgements

The authors would like to thank Kristina Lövgren, Lilian Karlsson, Ann Wikström and Ulric Pedersen for their technical assistance.

ID: nareshrao Time: 14:12 I Path: N:/3b2/IJC#/Vol00000/130510/APPFile/JW-IJC#130510

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA: a cancer journal for clinicians* 2011; 61:69–90.
2. Parris TZ, Danielsson A, Nemes S, et al. Clinical implications of gene dosage and gene expression patterns in diploid breast carcinoma. *Clin Cancer Res* 2010;16:3860–74.
3. McCarty KS, Jr, Miller LS, Cox EB, et al. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch Pathol Lab Med* 1985;109:716–21.
4. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004;10:7252–9.
5. Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res* 2012;18:4465–72.
6. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res* 2011;17:6012–20.
7. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007;356:217–26.
8. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25: 1239–46.
9. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351: 2817–26.
10. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160–7.
11. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262–72.
12. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
13. Myal Y, Gregory C, Wang H, et al. The gene for prolactin-inducible protein (PIP), uniquely expressed in exocrine organs, maps to chromosome 7. *Somat Cell Mol Genet* 1989;15:265–70.
14. Burgi W, Schmid K. Preparation and properties of Zn-alpha 2-glycoprotein of normal human plasma. *J Biolo Chem* 1961;236:1066–74.
15. Hale LP, Price DT, Sanchez LM, et al. Zinc alpha-2-glycoprotein is expressed by malignant prostatic epithelium and may serve as a potential serum marker for prostate cancer. *Clin Cancer Res* 2001;7:846–53.
16. Tada T, Ohkubo I, Niwa M, et al. Immunohistochemical localization of Zn-alpha 2-glycoprotein in normal human tissues. *J Histochem Cytochem* 1991;39:1221–6.
17. Todorov PT, McDevitt TM, Meyer DJ, et al. Purification and characterization of a tumor lipid-mobilizing factor. *Cancer Res* 1998;58:2353–8.
18. Bing C, Bao Y, Jenkins J, et al. Zinc-alpha2-glycoprotein, a lipid mobilizing factor, is expressed in adipocytes and is up-regulated in mice with cancer cachexia. *Proc Natl Acad Sci U S A* 2004;101: 2500–5.
19. Sanchez LM, Vizoso F, Diez-Itza I, et al. Identification of the major protein components in breast secretions from women with benign and malignant breast diseases. *Cancer Res* 1992;52:95–100.
20. Abdul-Rahman PS, Lim BK, Hashim OH. Expression of high-abundance proteins in sera of patients with endometrial and cervical cancers: analysis using 2-DE with silver staining and lectin detection methods. *Electrophoresis* 2007;28:1989–96.
21. Brysk MM, Lei G, Adler-Storthz K, et al. Zinc-alpha2-glycoprotein expression as a marker of differentiation in human oral tumors. *Cancer Lett* 1999;137:117–20.
22. Freije JP, Fueyo A, Uria J, et al. Human Zn-alpha 2-glycoprotein cDNA cloning and expression analysis in benign and malignant breast tissues. *FEBS Lett* 1991;290:247–9.
23. Gagnon S, Tetu B, Dube JY, et al. Expression of Zn-alpha 2-glycoprotein and PSP-94 in prostatic adenocarcinoma. *An immunohistochemical study of 88 cases. Am J Pathol* 1990;136:1147–52.
24. Irmak S, Tilki D, Heukeshoven J, et al. Stage-dependent increase of orosomucoid and zinc-alpha2-glycoprotein in urinary bladder cancer. *Proteomics* 2005;5:4296–304.
25. Lei G, Arany I, Selvanayagam P, et al. Detection and cloning of epidermal zinc-alpha 2-glycoprotein cDNA and expression in normal human skin and in tumors. *J Cell Biochem* 1997;67:216–22.
26. Zhu HJ, Dong CX, Pan H, et al. rs4215 SNP in zinc-alpha2-glycoprotein gene is associated with obesity in Chinese north Han population. *Gene* 2012;500:211–5.
27. Bao Y, Bing C, Hunter L, et al. Zinc-alpha2-glycoprotein, a lipid mobilizing factor, is expressed and secreted by human (SGBS) adipocytes. *FEBS Lett* 2005;579:41–7.
28. He N, Brysk H, Tyring SK, et al. Zinc-alpha(2)-glycoprotein hinders cell proliferation and reduces cdc2 expression. *J Cell Biochem* 2001; Suppl 36:162–9.
29. Kong B, Michalski CW, Hong X, et al. AZGP1 is a tumor suppressor in pancreatic cancer inducing mesenchymal-to-epithelial transdifferentiation by inhibiting TGF-beta-mediated ERK signaling. *Oncogene* 2010;29:5146–58.
30. Mazoujian G, Bodian C, Haagensen DE, Jr, et al. Expression of GCDFP-15 in breast carcinomas. *Relationship to pathologic and clinical factors. Cancer* 1989;63:2156–61.
31. Debily MA, Marhomy SE, Boulanger V, et al. A functional and regulatory network associated with PIP expression in human breast cancer. *PLoS One* 2009;4:e4696.
32. Haagensen DE, Stewart P, Dilley WG, et al. Secretion of breast gross cystic disease fluid proteins by T47D breast cancer cells in culture—modulation by steroid hormones. *Breast Cancer Res Treat* 1992;23:77–86.
33. Yadav VK, Kumar V, Chhikara N, et al. Purification and characterization of a native zinc-binding high molecular weight multiprotein complex from human seminal plasma. *J Separ Sci* 2011;34: 1076–83.
34. Baniwal SK, Little GH, Chimge NO, et al. Runx2 controls a feed-forward loop between androgen and prolactin-induced protein (PIP) in stimulating T47D cell proliferation. *J Cell Physiol* 2012; 227:2276–82.
35. Jones MD, Liu JC, Barthel TK, et al. A proteasome inhibitor, bortezomib, inhibits breast cancer growth and reduces osteolysis by downregulating metastatic genes. *Clin Cancer Res* 2010;16:4978–89.
36. Clark JW, Snell L, Shiu RP, et al. The potential role for prolactin-inducible protein (PIP) as a marker of human breast cancer micrometastasis. *Br J Cancer* 1999;81:1002–8.
37. Pagani A, Sapino A, Eusebi V, et al. PIP/GCDFP-15 gene expression and apocrine differentiation in carcinomas of the breast. *Virch Archiv: Int J Pathol* 1994;425:459–65.
38. Selim AA, El-Ayat G, Wells CA. Immunohistochemical localization of gross cystic disease fluid protein-15, -24 and -44 in ductal carcinoma in situ of the breast: relationship to the degree of differentiation. *Histopathology* 2001;39:198–202.
39. Cross SS, Hamdy FC, Deloulme JC, et al. Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers. *Histopathology* 2005;46:256–69.
40. Odink K, Cerletti N, Bruggen J, et al. Two calcium-binding proteins in infiltrate macrophages of rheumatoid arthritis. *Nature* 1987;330: 80–2.
41. Thorey IS, Roth J, Regenbogen J, et al. The Ca^{2+}-binding proteins S100A8 and S100A9 are encoded by novel injury-regulated genes. *J Biol Chem* 2001;276:35818–25.
42. Ghafouri B, Tagesson C, Lindahl M. Mapping of proteins in human saliva using two-dimensional gel electrophoresis and peptide mass fingerprinting. *Proteomics* 2003;3:1003–15.
43. McKiernan E, McDermott EW, Evoy D, Crown J, Duffy MJ. The role of S100 genes in breast cancer progression. *Tumour Biol* 2011;32:441–50.
44. Hashemi FB, Mollenhauer J, Madsen LD, et al. Myeloid-related protein (MRP)-8 from cervicovaginal secretions activates HIV replication. *AIDS (London, England)* 2001;15:441–9.
45. Moon A, Yong HY, Song JI, et al. Global gene expression profiling unveils S100A8/A9 as candidate markers in H-ras-mediated human breast epithelial cell invasion. *Mol Cancer Res* 2008;6: 1544–53.
46. Townsley FM, Aristarkhov A, Beck S, et al. Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. *Proc Natl Acad Sci U S A* 1997;94:2362–7.
47. Berlingieri MT, Pallante P, Sboner A, et al. UbcH10 is overexpressed in malignant breast carcinomas. *Eur J Cancer* 2007;43:2729–35.
48. Okamoto Y, Ozaki T, Miyazaki K, et al. UbcH10 is the cancer-related E2 ubiquitin-conjugating enzyme. *Cancer Res* 2003;63:4167–73.
49. Lee JJ, Foukakis T, Hashemi J, et al. Molecular cytogenetic profiles of novel and established human anaplastic thyroid carcinoma models. *Thyroid* 2007;17:289–301.
50. Takahashi Y, Ishii Y, Nishida Y, et al. Detection of aberrations of ubiquitin-conjugating enzyme E2C gene (UBE2C) in advanced colon cancer with liver metastases by DNA microarray and two-color FISH. *Cancer Genet Cytogenet* 2006; 168:30–5.

Supporting information Table S1. Multigene signatures for breast carcinoma

Assay description	Biomarkers (n)	Method and sample conditions	Clinical use	Biomarkers found in the signature in common with the 13-marker signature *
MAMMAPRINT® (1)				
FDA approved assay for pN0 breast cancer patients of all ages with tumors less than 5 cm and either ER- or ER+ to predict risk for metastasis and determine which patients will benefit from chemotherapy	70	Microarray, fresh/frozen or formalin-fixed tissue	Tamoxifen, adjuvant and neoadjuvant chemotherapy	***SCUBE2, STK32B***
CLINICAL TREATMENT SCORE				
An algorithm which consisting of axillary lymph node status, tumor size, histological grade, age, and treatment	N/A	N/A		N/A
IHC4 (2)				
A four-marker signature (ER, PgR, HER2, Ki67) which provides independent prognostic information after adjusting for established clinical variables	4	IHC, FFPE		None
ONCOTYPE Dx™ RECURRENCE SCORE (3)				
Signature used to calculate a disease recurrence score in early-stage ER-positive breast cancer	21	qPCR, FFPE	Tamoxifen, Adjuvant, CMF	***SCUBE2***
PREDICTOR ANALYSIS OF MICROARRAY (PAM50) (4, 5)				
Signature to classify breast tumors into the intrinsic molecular subtypes	50	qPCR, fresh/frozen	Neo-adjuvant chemotherapy	***UBE2C***
BREAST CANCER INDEX (6)				
Signature combining the molecular grade index and *HOXB13:IL17BR* to identify a subgroup of early-stage ER-positive breast cancer patients with an unfavorable prognosis despite adjuvant endocrine therapy	5	qPCR, FFPE		None
ENDOPREDICT (7)				
Signature to predict distant recurrence in ER-positive, HER2/*neu*-negative breast cancer treated with adjuvant endocrine therapy	8	qPCR		***AZGP1, STC2, UBE2C***
GENOMIC GRADE INDEX (8, 9)				
Signature to define histological grade (high or low genomic grade) in estrogen receptor-positive breast cancer using molecular profiling.	97	Microarray, fresh/frozen or formalin-fixed tissue		***NME5, UBE2C***
ROTTERDAM 76-GENE SIGNATURE (10)				
Signature to predict distant recurrence in axillary lymph node-negative breast cancer patients	76	Microarray, fresh/frozen or formalin-fixed tissue		None
WOUND RESPONSE SIGNATURE (11)				
Signature depicting a wound healing response in fibroblasts from ten anatomical regions after serum exposure	446	Microarray, fresh/frozen or formalin-fixed tissue		
INVASIVENESS GENE SIGNATURE (12)				
Signature that can differentiate highly tumorigenic CD44+CD24-/low cells from normal breast epithelium	186	Microarray, fresh/frozen or formalin-fixed tissue		***STC2***

*Note: Gene targets discussed in this article are displayed in bold text.

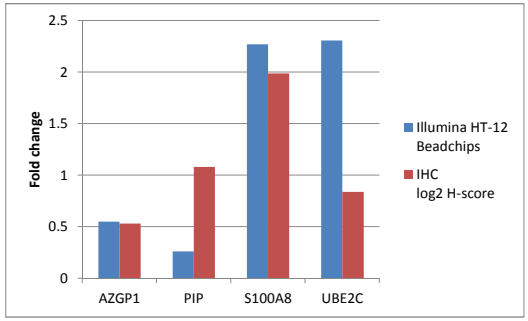
Abbreviations: CMF = Cyclophosphamide, methotrexate and 5-fluorouracil; FFPE = Formalin-fixed, paraffin-embedded; IHC = Immunohistochemistry; qPCR = Quantitative real-time PCR; N/A = Not applicable.

References

1. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415:530-6.
2. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol. 2011;29:4273-8.
3. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351:2817-26.
4. Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. Clin Cancer Res. 2012;18:4465-72.
5. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160-7.
6. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res. 2008;14:2601-8.
7. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17:6012-20.
8. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol. 2007;25:1239-46.
9. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 2006;98:262-72.
10. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 2005;365:671-9.
11. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS biology. 2004;2:E7.
12. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med. 2007;356:217-26.

Supporting information Table S2: Correlation of IHC protein expression and Illumina HumanHT-12 microarray gene expression analyses

	Relative expression (fold change)		
Gene Symbol	Illumina HT-12 Beadchips	IHC log2 H-score	Correlation
AZGP1	0.55	0.53	0.50732694
PIP	0.26	1.08	
S100A8	2.27	1.99	
UBE2C	2.31	0.84	



	Relative expression (fold change)		
Gene Symbol	Illumina HT-12 Beadchips	IHC log2 H-score	Correlation
CBX2	1.89	0.82	-0.31908981
DNALI1	0.49	1.43	
NME5	0.57	0.54	
SCUBE2	0.44	0.93	
SERPINA11	0.75	1.23	
STC2	0.4	3.71	
SUSD3	0.44	1.09