



Published in final edited form as:

*Clin Cancer Res.* 2016 October 15; 22(20): 5068–5078. doi:10.1158/1078-0432.CCR-16-0171.

## High intra-tumoral stromal content defines Reactive breast cancer as a low-risk breast cancer subtype

Jennifer B. Dennison<sup>1</sup>, Maria Shahmoradgoli<sup>1</sup>, Wenbin Liu<sup>2</sup>, Zhenlin Ju<sup>2</sup>, Funda Meric-Bernstam<sup>3</sup>, Charles M. Perou<sup>4</sup>, Aysegul A. Sahin<sup>5</sup>, Alana Welm<sup>6</sup>, Steffi Oesterreich<sup>7</sup>, Matthew J. Sikora<sup>7</sup>, Robert E. Brown<sup>8</sup>, and Gordon B. Mills<sup>1,\*</sup>

<sup>1</sup>Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

<sup>2</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

<sup>3</sup>Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

<sup>4</sup>Department of Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

<sup>5</sup>Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

<sup>6</sup> Huntsman Cancer Institute, University of Utah, Salt Lake City UT 84112

<sup>7</sup>Department of Pharmacology and Chemical Biology, University of Pittsburgh Cancer Institute, Pittsburgh, PA 15213 USA.

<sup>8</sup>Department of Pathology and Laboratory Medicine, The University of Texas Health Science Center, Houston, TX 77030 USA.

### Abstract

**Purpose**—The current study evaluated associative effects of breast cancer cells with the tumor microenvironment and its influence on tumor behavior.

**Experimental design**—Formalin-fixed paraffin embedded tissue and matched protein lysates were evaluated from two independent breast cancer patient data sets (TCGA and MD Anderson). Reverse-phase protein arrays (RPPA) were utilized to create a proteomics signature to define

\* **Corresponding author:** Gordon B. Mills, MD, PhD. Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. gmills@mdanderson.org Tel. (713) 563-4200.

**Potential Conflicts of Interest:** none

#### Author's Contributions

Concept and design: J.B. Dennison, G.B. Mills, M. Shahmoradgoli, F. Meric-Bernstam, and C.M. Perou.

Design of experiments and/or acquisition of data: J.B. Dennison, G.B. Mills, A.A. Sahin, R.E. Brown

Provided materials: A. Welm, S. Oesterreich, and M.J. Sikora

Analysis and interpretation of data: J.B. Dennison, W. Liu, Z. Ju

Writing, review, and/or revision of manuscript: J.B. Dennison, W. Liu, Z. Ju, M. Shahmoradgoli, F. Meric-Bernstam, A.A. Sahin, R.E. Brown, C.M. Perou, A. Welm, S. Oesterreich, M.J. Sikora, and G.B. Mills

breast tumor subtypes. Expression patterns of cell lines and normal breast tissues were utilized to determine markers that were differentially expressed in stroma and cancer cells. Protein localization and stromal contents were evaluated for matched cases by imaging.

**Results**—A subtype of breast cancers designated “Reactive,” previously identified by RPPA that was not predicted by mRNA profiling, was extensively characterized. These tumors were primarily estrogen receptor (ER)-positive/human epidermal growth factor receptor (HER)2-negative, low-risk cancers as determined by enrichment of low-grade nuclei, lobular or tubular histopathology, and the luminal A subtype by PAM50. Reactive breast cancers contained high numbers of stromal cells and the highest extracellular matrix content typically without infiltration of immune cells. For ER-positive/HER2-negative cancers, the Reactive classification predicted favorable clinical outcomes in the TCGA cohort (HR = 0.36,  $P < 0.05$ ).

**Conclusions**—A protein stromal signature in breast cancers is associated with a highly differentiated phenotype. The stromal compartment content and proteins are an extended phenotype not predicted by mRNA expression that could be utilized to sub-classify ER-positive/HER2-negative breast cancers.

### Keywords

intra-tumoral stroma; reactive; breast cancer classification; lobular; overtreatment

---

### Introduction

Molecular stratification by mRNA has identified multiple subtypes of breast cancers that can predict outcomes independently of clinical characteristics and standard protein biomarkers (1,2). These molecular subtypes of cancer were reported to be associated with tumor microenvironments of different tumor and stromal cell contents and expression of extracellular matrix mRNA and proteins (3,4). In fact, rigorous morphological evaluation of the tumor microenvironment was reported to provide similar prognostic information content relative to the cancer epithelium itself (5). However, the criteria used to evaluate the tumor microenvironment are not well established and thus are not universally applied, particular in the clinical management of tumors. Also, it is unclear mechanistically how characteristics of the tumor microenvironment influence patient outcomes. Some reports with multiple malignancies including lung, colon, and gastric cancer suggested that higher quantities of intra-tumoral stromal and myofibroblast contents were associated with worse clinical outcomes and metastasis (6-8). In breast cancer, conflicting imaging studies have been published on stromal content and prognosis. For ER-positive breast cancers, improved outcomes were associated with increased stromal content (9). In contrast, a high stromal content predicted poor survival in triple-negative breast cancers (10-13). These clinical studies of breast cancer microenvironment, with outcome as the primary endpoint, may have confounding results because of the presence of multiple molecular subtypes of epithelium, different clinical standard-of-care treatments, and different patient selection criteria.

Previously, we reported an RPPA subtype called “Reactive” breast cancer with high expression of extracellular matrix proteins (14,15). Herein we demonstrate that the Reactive subtype represents a more differentiated and relatively better-behaved disease, enriched in

lobular and luminal A breast cancers. Our data suggest that a high stromal protein signature marks fundamental biological differences between breast cancers and may help identify breast cancers with excellent prognosis.

## Patients and Methods

### Clinical samples and cell lines

Two independent sets of breast cancers with matched protein lysates and formalin-fixed, paraffin-embedded (FFPE) sections were utilized: TCGA ( $n = 885$ ) and MDACC ( $n = 63$ ). For the TCGA cohort, only a subset of cases with RPPA data was available with matched unstained FFPE sections ( $n = 109$ ).

Breast cancer cell lines (T47D, MCF7) were obtained from the MD Anderson Characterized Cell Line Core facility (Houston, TX). The identities of all cell lines were verified using AmpF/STR Identifier kit (Applied Biosystems) by the core facility and used within 6 months of obtaining the cell lines. The lines were cultured in DMEM supplemented with 5% fetal bovine serum (FBS) at 37°C in 5% carbon dioxide atmosphere. Cell lines were routinely tested for *Mycoplasma* infection using a MycoTect Kit (Invitrogen).

MDA-MB-134 (ATCC) and Sum44PE (Asterand) cells were cultured as previously described (16). The cell lines were authenticated annually by PCR RFLP analyses and confirmed to be *Mycoplasma* negative at the University of Pittsburgh Cell Culture and Cytogenetics Facility (Pittsburgh, PA).

Cancer tissues from the University of Utah (A.W.) were mastectomy specimens from patients without prior therapy. Normal specimens were collected from matched grossly uninvolved tissues. Tissues were collected and de-identified using approved IRB#10924.

### Immunohistochemistry and imaging studies

Board-certified pathologists blinded to the RPPA results performed the histopathology classification and the manual scoring of sections. Manual scoring by pathologists for the TCGA cohort was performed using  $\alpha$ -SMA (R.E.B.) and H&E stained sections (TCGA pathology team). For the H&E scored cases of the TCGA, epithelial content was reported as follows: proportion of epithelium in invasive portion by area (excluding areas of necrosis): <25% (Low; score 1), 25-75% (Moderate; score 2), >75% (High; score 3). To obtain scores for the intra-tumoral stroma (ITS) of the TCGA cases by H&E, the epithelial content values were converted to stromal content; score 1 was converted to score 3, and score 2 was unchanged. Nottingham grade for the TCGA cases was a combined score of epithelial tubule formation, mitosis, and nuclear pleomorphism (3 each for a maximum total score of 9, score was separated into three bins: [1] 3,4; [2] 5,6,7; [3] 8,9).

The stromal content using H&E stained sections (MDACC cohort) or  $\alpha$ -SMA stained sections (TCGA cohort) were quantified by imaging (InForm 2.0) similar to methods previously described (17). FFPE tumor sections were first scanned at low power (40x) to identify tissue and then high-powered field (HPF) images ( $n = 50$  maximum) were collected randomly across the whole section at 200x magnification (Vectra 2.0, Perkin Elmer). All

fields of invasive cancer without edge effects, normal tissue, or necrosis were selected for scoring by an investigator blinded to the identities of the cases.

Unstained sections from selected matched TCGA cases were used for immunohistochemistry staining of various proteins. Slides were dried at 60°C, deparaffinized in xylene, quenched in 3% hydrogen peroxide/methanol for 10 min, and rehydrated through sequentially graded ethanol. Antigen retrieval was performed by EDTA incubation (1 mmol/L, pH 8.0) using a pressure cooker at 125°C/25 psi for 5 min. Slides were cooled for 20 min, rinsed with distilled water, and washed in Tris-buffered saline. Using a DAKO Autostainer (DAKO, Carpinteria, CA), the slides were blocked for 30 min in horse serum and incubated with primary antibody for 1 h. Primary antibodies were matched to the RPPA platform: Caveolin-1 (1:200), E-cadherin (1:200), and RBM15 (1:1000). The Universal ABC Elite kit (Vectastain, Burlingame, CA) with 3,3'-diaminobenzidine development was used to visualize antibody binding, and the slides were counterstained with hematoxylin. For  $\alpha$ -SMA staining, slides were incubated for 20 min at room temperature with primary antibody, Dako M0851 (mouse, clone 1A4), 1:500 dilution and developed using polymer-HRP and 3,3'-diaminobenzidine per the manufacturer's instructions (PowerVision+, Leica PV6104). Breast cancer tissues without primary antibody incubation were utilized as negative controls.

### Reverse-phase protein array (RPPA)

Lysates from flash-frozen tissues from the TCGA and MDACC cohorts were prepared and analyzed by RPPA as described previously (14). For the MDACC cohort, H&E sections were obtained from each flash frozen tissue. Using these sections, a pathologist (A.S.) verified the flash frozen tissue was invasive carcinoma prior to RPPA protein lysate preparation of the frozen tissue.

### Analytical and statistical methods

mRNA and protein expression differences between breast cancers and cell lines were assessed using Student's t-tests and one-way ANOVA. Survival curves were determined using the Kaplan-Meier method, and differences between groups were evaluated using the Log-rank test. The association between sample cluster (Cluster II and Other) and clinical factors was detected using chi-square test. Statistical analyses were performed using Prism 5.0c (GraphPad Software, La Jolla, CA) or R, version 2.15.1 (<http://www.r-project.org/>). Consensus clustering was performed using R, and heatmaps were generated using an R-package NGCHM (see Supplementary materials).  $P < 0.05$  (two-sided) was considered significant.

## Results

### Identification of Reactive breast cancer subtype by RPPA

RPPA analyses were performed on breast cancer protein lysates from two independent cohorts from the TCGA and the University of Texas MD Anderson Cancer Center (MDACC cohort). For the TCGA cohort, a subset was previously evaluated as an individual cohort ( $n = 403$ ) (14) and included as part of a combined pan-cancer analysis ( $n = 747$ ) (15); however,

the current set is expanded ( $n = 885$ ) and analyzed independently of other malignancies. Four groups were evident by consensus clustering including two primarily ER-positive, one HER2-positive, and one basal-like group (Supplementary Fig. S1). To evaluate the key features of the RPPA subtypes, a small percentage of the TCGA cases without strong statistical association to any cluster were removed to generate a core group of TCGA cases ( $n = 774$ ) (Fig. 1A, see Supplementary Materials).

Previously, we reported Cluster II (Fig. 1A, Supplementary Fig. S1) of the TCGA cohort as two separate “Reactive” breast cancer subtypes (14). These two groups within Cluster II were still evident in the entire cohort (Supplementary Fig. S1) and to a lesser extent in the core cohort (Fig. 1A). However, because these two groups were defined by a single set of markers with a graded expression, we evaluated them as one subtype termed “Reactive” for this manuscript and recommend that they be considered as a single group in future studies. Prediction models using RPPA protein expression data were highly effective in identifying Cluster II (Reactive) breast cancers for all core cases and the ER-positive/HER2-negative group (Supplementary Tables 1 and 2, see Supplementary Materials). For the ER-positive/HER2-negative group, key determinants positively associated were annexin-1, collagen VI, and HSP70 and negatively associated were BAP1, phosphomyosinIIA, and PRDX1 with more modest contributions from phosphoERalpha and GATA3.

Reactive (Cluster II) cases were mostly ER-positive (82%) and were distinctive from other ER-positive cases in Cluster I (99% ER-pos, Fig. 1A). The Reactive subtype was also enriched in lobular cancers and thus *CDH1* mutations and low-risk characteristics of breast cancers including low Nottingham grade, luminal A breast cancers, and wild-type *TP53* (Table 1). Importantly, this enrichment of characteristics for Reactive breast cancers was still maintained within the ER-positive/HER2-negative clinical subtype (Table 1). These data suggest that the Reactive breast cancer subtype is independent of the existing clinical protein biomarkers (ER and HER2).

By RPPA analysis, Reactive breast cancer was a distinctive group of cancers with relatively high expression of stromal proteins such as collagen-VI. Surprisingly, the cancers expressed lower levels of proteins associated with ER-positive breast cancer such as ER-alpha and GATA3 as compared to Cluster I (Fig. 1B, Supplementary Table 3, Supplementary Fig. S2) even though the Reactive cancers were ER-alpha positive as determined by clinical criteria. Thus, we hypothesized that epithelial-derived proteins such as ER-alpha were diluted by extracellular matrix (ECM) proteins or by increased numbers of stromal cells.

To test this hypothesis, localization and relative expression of selected proteins by immunohistochemistry were evaluated for representative IDC (invasive ductal carcinoma) cases using the same antibodies as the RPPA assays: caveolin-1 (high in Reactive group, Fig. 1A and 1B), E-cadherin (low in Reactive group, Fig. 1A), and RBM15 (low in Reactive group, Fig. 1A and 1B). For all three proteins, the expression patterns and intensities were consistent with a higher amount of stromal cells and/or ECM diluting the epithelium in Reactive breast cancers (Supplementary Fig. S3 to S5). Caveolin-1 was exclusively expressed in the stromal compartment for luminal and Reactive cancers (Supplementary Fig. S3). E-cadherin was expressed by epithelium in the cell membrane of all IDC cases without

an apparent difference in intensity in the tumor cells between luminal and Reactive breast cancers (Supplementary Fig. S4). A decreased number of cancer cells with an increase in ECM could contribute to low total lysate levels of RBM15 because the protein was equally expressed and localized to the nucleus of all cells independently of RPPA subtype (Supplementary Fig. S5). In support of an epithelial protein dilution effect, mRNA profiling approaches like the PAM50 have not previously associated the RPPA Reactive subtype with high stromal marker expression. In our study, the mRNA expression levels of caveolin-1, collagen VI, ER-alpha, and RBM15 were highly variable between the RPPA subtypes (Supplementary Fig. S6A) and not predictive of protein content especially within the Cluster II/Reactive subtype (Supplementary Fig. S6B).

To further confirm that Reactive cases were enriched in stromal proteins, the RPPA proteins on the platform were first classified as primarily “stroma” or ER-positive breast “cancer” derived. To classify the proteins, expression patterns of normal breast tissues were compared to those of ER-positive breast cancer cell lines including representative IDC and invasive lobular carcinoma (ILC) cell lines (Supplementary Fig. S7, Supplementary Table 4). Normal breast tissues were chosen as the stromal-enriched tissue because at least 70% of non-fatty areas in normal tissue are composed of stroma rather than epithelium by imaging (18). The results of IDC and ILC comparisons to normal breast tissue were aligned with two exceptions; cell junction proteins, E-cadherin and  $\beta$ -catenin, were characteristically low in lobular breast cancer. Using these protein classifications, RPPA expression patterns of Cluster II were confirmed to be highly associated with stromal proteins and reduced expression of cancer cell proteins for the TCGA cohort (Supplementary Fig. S8).

### Patient outcomes for Reactive breast cancers

To evaluate the clinical impact of cancers with high expression of stromal proteins, patient outcomes were evaluated within the RPPA subtypes; Reactive breast cancers had the best outcomes while the basal subtype had the worst clinical outcomes (Fig. 1C). As expected when including triple-negative and HER2-positive cancers, poor outcomes were also highly associated with high-risk characteristics of breast cancers including tumor size, histopathology, Nottingham grade, lymph node status, and PAM50 classification (Table 2). However, within the ER-positive/HER2-negative clinical subtype, only young age and the RPPA Reactive breast cancer subtype predicted improved survival outcomes demonstrating the potential utility of RPPA subtyping within this highly relevant clinical subtype (Table 2). Interestingly, RPPA Reactive subtyping also stratified ILC risk in terms of outcomes although the total events for ILC cases were low for the TCGA cohort (log-rank P-value = 0.019; 0 events for Reactive,  $n = 77$ , versus 4 events for non-Reactive,  $n = 63$ ).

### Stromal content by quantitative imaging of breast cancers

To confirm that the Reactive breast cancers were higher in stromal content, matched cases from the TCGA were evaluated by imaging (Fig. 2). To perform a quantitative evaluation of the cases for stromal content including area and cell count, representative matched sections from the TCGA cohort were stained with  $\alpha$ -SMA to mark cells within the stromal compartment (myofibroblasts, carcinoma-associated fibroblasts). All available cases for ILC and representative IDC cases from each RPPA subtype of the TCGA cohort were stained and



evaluated. InForm Vectra software was utilized for compartment segmentation and cellular counting for multiple high-powered fields randomly selected across the sections (Fig. 2A, Supplementary Fig. S9). The stromal content both by area (Fig. 2B) and by cell count (Fig. 2C) were approximately 2-fold higher within the RPPA Reactive breast cancer subtypes as compared to luminal for both ILC and IDC cases. Consistent with the clustering results (Fig. 1A), the amount of stromal content within the RPPA subtypes was independent of ILC or IDC classification (Fig. 2B and 2C). These results also demonstrated that RPPA basal-like breast cancers contained the lowest quantity of stroma both by area and cell counting (Fig. 2B and 2C).

To validate the findings from the TCGA cohort, cases from an independent MDACC cohort of ER-positive/HER2-negative cancers were similarly evaluated by RPPA and by imaging. Like that observed for the TCGA cohort, consensus clustering of MDACC cases revealed two primary clusters including a subtype enriched in stromal proteins, Reactive breast cancer (Supplementary Fig. S10A and 10B, Supplementary Table 3). The differences in ER protein expression between the two subtypes were evaluated by comparing the % positivity of ER by IHC (Supplementary Fig. S10C). While the %ER positivity was statistically lower in the Reactive subtype, this difference was caused primarily by 5 cases with the lowest %ER expression (<50%); most Reactive cancers were highly ER-positive (80 to 100%). Higher stromal content by area in the Reactive subtype was confirmed by imaging in the ER-positive/HER2-negative MDACC cohort (Supplementary Fig. S10D). The highest stromal contents were also associated with normal breast, ILC, and lower nuclear grade cases (Supplementary Fig. S10D).

To evaluate whether there were qualitative differences in the stromal compartment, we calculated the cell density within the cancer cell and stromal compartments for the TCGA cohort (Supplementary Fig. S11A). As expected, the stromal compartment cell density was much less cell-enriched as compared to the cancer compartment (25% of the cancer compartment on average, Supplementary Fig. S11A). While no differences were observed within the cancer compartment (Supplementary Fig. S11B), the stromal cell density was higher in the Reactive breast cancer subtype of ILC cases but not for IDC cases (Fig. 2D). However, the stromal compartments of the basal-like IDC cases were on average less stromal-cell enriched than the luminal IDC cases (Fig. 2D). These data suggest that the stromal cell concentration is different between RPPA subtypes, not just the amount of ECM, although stromal cell concentration and ECM may be associated with one another.

To further characterize the stroma, we also evaluated myofibroblast content by quantifying the staining intensity of  $\alpha$ -SMA. Interestingly, the highest  $\alpha$ -SMA staining (3+) was observed for some of the ILC cases within the Reactive subtype (Supplementary Fig. S11C, S11D, S11E, and S12). However, consistent with previous studies (19), there was substantial variability in SMA staining (Supplementary Fig. S11C). No differences in  $\alpha$ -SMA staining were detected between RPPA subtypes for the IDC cases (Supplementary Figs. S11C and S13).

### mRNA/DNA signatures of stromal characteristics for Reactive breast cancer

Given that a higher density and the effective overall number of stromal cells were increased in the RPPA Reactive subtype, we hypothesized that Reactive breast cancer may have altered mRNA and/or DNA cell-based scores (ABSOLUTE, ESTIMATE) based on approaches used to evaluate stromal content and cell types including purity measures, proliferation, and immune cell contents as defined previously (4,15,20,21). We evaluated the association of these scores to RPPA subtyping of breast cancer within histopathology groups (ILC and IDC, Supplementary Fig. S14). For IDC and sometimes ILC cases, Reactive breast cancers were statistically less pure and contained higher stromal content using DNA- or mRNA-derived scores (Supplementary Fig. S14A, S14B, S14C, S14D, and S14E). However, purity and stromal content scores were not sufficient to identify Reactive breast cancers.

Furthermore, the differences between the luminal and Reactive breast cancer subtypes were minor (7% difference for ABSOLUTE scores, Supplemental Supplementary Fig. S14A), and the standard deviations of the scores were high between subtypes (ESTIMATE scores, Supplemental Supplementary Fig. S14B).

Because high epithelial mesenchymal transition (EMT) scores may suggest a higher mesenchymal cell content, mRNA-derived EMT scores (22) were evaluated for the RPPA subtypes. As expected for cancers with a high fibroblast content, Reactive breast cancers had the highest EMT scores as compared to all other subtypes including luminal, HER2, and basal subtypes (Supplementary Fig. S14D). These associations were independent of histopathology demonstrating that loss of E-cadherin in ILC cases did not explain the high EMT scores. This result is also consistent with Reactive breast cancers as highly differentiated cancers, on the opposite end relative to the “claudin-low” cancers, a rare subtype in which the cancer cells themselves have EMT characteristics including loss of E-cadherin (23).

We also evaluated immune infiltration for the Reactive subtype. Consistent with previous reports of tumor-infiltrating lymphocytes in triple-negative breast cancer (24), the RPPA subtype with the highest immune score was Cluster IV, which largely corresponds to PAM50 basal-like breast cancers (Supplementary Fig. S14E and S14F). No differences between luminal and Reactive breast cancer subtypes were observed in immune response by immune response score (Supplementary Fig. S14E). Stromal lymphocytes were rare in ILC breast cancer in particular; only 8 events of immune infiltration were recorded for all the scored ILC cases (N = 73, Supplementary Fig. S14F).

Thus, the data suggest that mRNA scores for tumor purity, stromal content, or EMT trend in the expected direction but cannot effectively identify the RPPA-based Reactive breast cancer subtype. Also, Reactive breast cancers are not characterized by immune infiltration.

### Intra-tumoral stromal (ITS) scoring to characterize breast cancers

In addition to our quantitative imaging evaluation of stromal content for selected cases, the stromal content could also be estimated by using intra-tumoral stromal (ITS) scores for almost the entire TCGA cohort. For matched cases, software quantification values of stromal areas were in good agreement with the manual scoring definitions (Fig. 2E, Spearman



correlation coefficient = 0.53,  $P = 0.001$ ), but with these established limits, there were relatively few high “3” (>75% stroma by area) and low scores “1” (<25% stroma by area). Consistent with the software quantification, manual scoring of the ITS content demonstrated within IDC that RPPA Reactive breast cancers contained lower amounts of epithelium in the invasive part of the tumor (Fig. 2F). Although the trend was in the expected direction, the amounts of epithelium in ILC cases were not statistically different between the luminal and Reactive RPPA subtypes (Fig. 2F).

### Morphology of IDC Reactive breast cancers

We used ITS scores to also understand the effects of histopathology and other clinical characteristics on the tumor microenvironment in the TCGA cohort. ER-pos/HER2-neg ILC contained higher quantities of stroma in the invasive part of the tumor as compared to IDC independently of ER status (Fig. 3A). Independent of histopathology, stromal contents of ER-pos/HER2-neg breast cancers were the lowest in cases predicted to be aggressive by nuclear scoring (Fig. 3B). Interestingly, this association was only evident for ER-pos/HER2-neg tumors (not triple-negative or PAM50-defined basal-like cancers, data not shown).

In conjunction with high stromal content, evaluation of Reactive breast cancer morphology by imaging revealed that IDC Reactive breast cancers were highly differentiated cancers often with neoplastic tubule formation as compared to RPPA luminal and basal-like cancers (Fig. 3C and 3D and Supplementary Fig. S13). These differences were quantified by pathological scoring of neoplastic tubules for the TCGA cohort (Fig. 3E). Importantly, the other two scores used in combination to calculate the overall nuclear score for IDC cases (Fig. 3B) were not associated, or only weakly associated, with stromal content (nuclear pleomorphism and mitotic figures, Supplementary Fig. S15). While these neoplastic tubules in Reactive breast cancer were clearly cancerous, as shown by loss of myoepithelium (Fig. 3C), these tumors were remarkably similar to normal breast tissue as shown by protein expression patterns by RPPA (Supplementary Fig. S16). The fact that normal breast tissues cluster with Reactive breast cancers by RPPA suggests that the cancer cells present in Reactive breast cancers do not substantially alter the protein composition of the stromal compartment, at least for the proteins included in the RPPA panel. The enrichment of PAM50 “normal-like” breast cancer within the Reactive subtype is also consistent with the normal, less aggressive behavior of Reactive breast cancers (Table 1).

### Discussion

A number of pathological criteria have been used to assess risk of relapse or metastasis in breast cancer, such as differentiation and Nottingham grade of cancer cells. However, corresponding histological effects on the tumor microenvironment and of the microenvironment on outcomes have not been well evaluated. The interaction of cancer cells and the tumor microenvironment is a promising area of research because new biomarkers and therapeutics targets may be discovered that are not present in the cancer cells; a recent example being the development and promise of immune checkpoint inhibitors.

The current study evaluated histological and proteomics markers of the tumor microenvironment in breast cancer by utilizing cases from the TCGA supplemented by a

sample set from the MD Anderson Cancer Center for confirmation. We discovered a subtype of breast cancers that was uniquely defined only by protein expression. These “Reactive” cancers were typically ER-positive, low-risk, and also more likely to have high expression of stromal proteins. The Reactive subtype compared to other breast cancers also had higher intra-tumoral stromal volumes and increased cell densities within the stromal compartment. Importantly, patients with ER-positive/HER2-negative disease and a Reactive subtype signature had an excellent prognosis and thus would not likely benefit from additional therapies.

Multiple RPPA studies including our own (14,15,25) have previously reported a group of breast cancers with apparently high expression of stromal proteins such as collagen and caveolin-1. Previously, the biological significance of these findings was not understood; investigators speculated that high caveolin-1 expression was related to loss of expression in the microenvironment of more aggressive cancers (25). Our study points to increased stromal content, not loss of caveolin-1 in the tumor microenvironment, as a major cause for this observation. In addition, the stroma itself had a higher cell density on average in Reactive breast cancers.

From our study, we now understand that the presence of high intra-tumoral stroma with highly differentiated cancer cells is a fundamental property of Reactive breast cancers. While it might be tempting to draw conclusions about the “protective” nature of the stromal compartment in the metastasis potential of Reactive breast cancers, clinical outcomes are confounded by fundamental differences in cancer cell biology (i.e. luminal A enrichment). We know that breast cancer cells with high intra-tumoral stromal content are slowly dividing, so a higher stromal content may simply be a result of increased ratios of replication rates for fibroblasts versus cancer cells. Consistent with this hypothesis, fast-growing animal tumor models including patient-derived xenograft models of breast cancer have low stromal content. From our study data, we cannot make any assessments on the role of myofibroblasts or stromal proteins such as suggested for collagen VI in the metastasis of breast cancers (26). However, we can say that the presence of high ECM content and/or a high number of stromal cells by itself (and not immune cells) did not increase the probability of relapse in breast cancer and further were associated with an improved outcome. These effects on outcomes appear to be contextual because opposite associations with outcomes have been reported for other cancers such as gastric cancer (6-8).

Strengths of the current work are the size and comprehensive molecular and pathological classifications available for the TCGA cohort. We were able to use case-matched protein array, gene array, and imaging data to demonstrate the fundamental differences between the Reactive subtype and other breast cancer tumors. We were also able to validate many of the key findings with an independent MDACC cohort. While this MDACC validation cohort had few events and was unable to verify the clinical outcomes endpoint, other clinical data previously published confirmed our findings of improved outcomes for ER-positive breast cancers with a high stromal content assessed by imaging (9).

As compared to our results for ER-positive cancers, the findings in the literature are contradictory for ER-negative breast cancers. Other investigators have reported that a high

stromal content by an mRNA signature or imaging is associated with resistance to chemotherapy and worse clinical outcomes (10-13,21). While these reports may seem in conflict with our results, breast cancers that are less “basal-like” would be expected to be more chemotherapy resistant and per our results, to have a higher stromal content. For triple-negative breast cancers, more late relapses might also be expected for less “basal-like” breast cancers like those observed for luminal breast cancer within the triple-negative clinical subtype, which can represent up to 25% of all triple-negative breast cancers (27). However, we were unable to confirm the results of other studies using data from the TCGA cohort; ER-negative/HER2-negative cancers with high stromal content by manual scoring were rare in the TCGA cohort (<4%). Also, Reactive subtyping did not predict clinical outcomes in ER-negative/HER2-negative disease (data not shown). Thus, our data does not support a clinical value for Reactive subtyping for triple-negative breast cancers.

Our study indicates that one possible clinical use for Reactive subtyping would be the sub-classification of ER-positive/HER2-negative breast cancers (Table 2). While we were unable to directly compare the results of Reactive subtyping to Oncotype-DX, we were able to show differences between RPPA and mRNA classification by PAM50. Other studies using mRNA to evaluate stromal content were consistent with our results; high collagen mRNA expression patterns were associated with ER-positive/low-grade breast cancers (3). However, we were unable to reproduce these results using the data from the TCGA cohort (data not shown). The variability in the stromal mRNA scoring of the TCGA cohort suggests that protein assays may more reliably classify breast cancers when it comes to assessing stromal features.

At this time, it is unclear what would be the most effective biomarker to identify Reactive breast cancers in the clinic. While RPPA profiling was effective in a highly controlled, translational study like the TCGA, we demonstrated that normal breast tissue and the Reactive subtype have similar protein profiles (Supplementary Fig. S16), and thus, contamination with normal tissue could reduce specificity in RPPA assays (or other protein assays) and cause many more breast cancers to artificially appear Reactive. Although IHC can easily distinguish normal tissues from cancers and circumvent this problem, the current study did not identify a Reactive IHC biomarker from the RPPA panel of proteins for the cancer or the stromal compartments. Although we presented some evidence (increased  $\alpha$ -SMA and stromal cell density) that the stroma may be distinctive, additional molecular studies are required to know whether the stroma itself is altered in Reactive cancers. In the current study, protein expression differences by RPPA appeared to be solely attributed to the intra-tumoral stromal content. Thus, our study suggests that an effective and practical biomarker to identify Reactive breast cancers in clinical practice may be intra-tumoral stromal content by imaging. In our imaging studies, we discovered that stromal content was bimodal, and a threshold of 48% by area could identify most Reactive breast cancers (Fig. 2B, 93% sensitivity, 67% specificity). However, using the thresholds selected by the TCGA pathology team for manual scoring, a 75% cutoff had poor sensitivity (20%), and a 25% cutoff had no specificity (0%) for predicting Reactive breast cancers. Interestingly, in another imaging study of ER-positive breast cancers, the optimal threshold to segregate cases by clinical outcomes was approximately 50% intra-tumoral stromal content by area (9). Based on our results and those of others (8), we conclude that a 50% threshold may have

prognostic value for clinical outcomes in ER-positive/HER2-negative breast cancers but requires testing in a prospective clinical trial. Another choice for histopathological classification of Reactive breast cancers could be neoplastic tubule formation in IDC, ER-pos/HER2-neg cases; we demonstrated a strong association of tubule formation with intra-tumoral stromal content within the TCGA cohort (Fig. 3E). However, intra-tumoral stromal content is a more universal criteria because many Reactive cancers are also ILC or ILC/IDC mixed cases where neoplastic tubule formation cannot predict patient outcomes.

In summary, the subtyping of breast cancer by high stromal protein content has the potential to identify tumors with the best clinical outcomes. This information could be useful in assessing risk of relapse, especially within the heterogeneous group of ER-positive/HER2-negative breast cancers. This study sets the stage for future studies that will be able to more accurately understand the biology of the disease and thus determine the most effective, personalized treatment options for all breast cancers.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank Pamela K. Johnston for help with the IHC assays and Jared Burks in the North Campus Flow Cytometry and Cellular Imaging Core Facility at MD Anderson. We also acknowledge the work of many investigators in the generation of data for the TCGA cohort, specifically for this manuscript, the TCGA Pathology Team including Andrew H. Beck, Yujing J. Heng, Susan C. Lester, Gary M.K. Tse, Rachel E. Factor, Laura C. Collins, Kimberly H. Allison, and Yunn-Yi Chen.

**Financial support:** This work was supported in part by the Susan G. Komen Foundation (KG081694); GlaxoSmithKline; American Cancer Society Clinical Research Professorship Grant (CRP-07-234); Entertainment Industry Foundation; Breast Cancer Research Foundation; and Susan G. Komen for the Cure Scientific Advisory Council Grant (SAC100013); and the National Institutes of Health through M. D. Anderson's Cancer Center Support Grant CA016672. J.B.D. was supported by a GlaxoSmithKline TRIUMPH post-doctoral fellowship and the American Cancer Society, Joe and Jessie Crump Medical Research Fund Postdoctoral Fellowship. Funding was also supported by the DoD Breast Cancer Research Program #W81XWH-12-1-0077 (A.L.W.), BCRF (S.O.), and a DoD BCRP Postdoctoral Fellowship BC11061 and NIH K99 CA193734 (M.J.S.). C.M.P. was supported by NCI Breast SPORE program (P50-CA58223-09A1), and R01-CA148761.

## References

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406:747–52. [PubMed: 10963602]
2. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98:10869–74. [PubMed: 11553815]
3. Bergamaschi, A.; Tagliabue, E.; Sørlie, T.; Naume, B.; Triulzi, T.; Orlandi, R., et al. *J Pathol*. Vol. 214. John Wiley & Sons, Ltd; 2008. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome.; p. 357-67.
4. Beck, AH.; Espinosa, I.; Gilks, CB.; van de Rijn, M.; West, RB. *Lab Invest*. Vol. 88. Nature Publishing Group; 2008. The fibromatosis signature defines a robust stromal response in breast carcinoma.; p. 591-601.
5. Beck, AH.; Sangoi, AR.; Leung, S.; Marinelli, RJ.; Nielsen, TO.; van de Vijver, MJ., et al. *Sci Transl Med*. Vol. 3. American Association for the Advancement of Science; 2011. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival.; p. 108ra113-3.

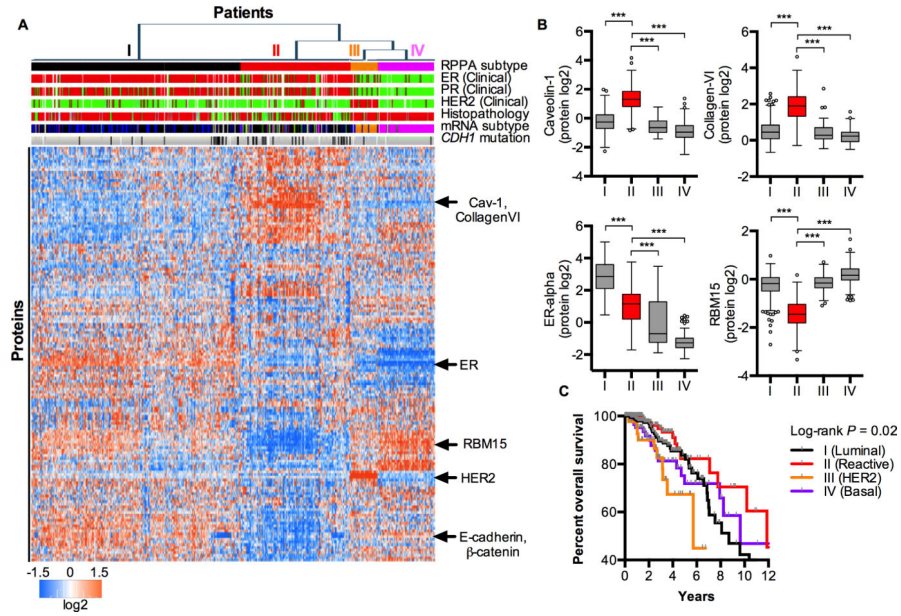
6. Wu, Y.; Grabsch, H.; Ivanova, T.; Tan, IB.; Murray, J.; Ooi, CH., et al. *Gut*. Vol. 62. BMJ Publishing Group Ltd and British Society of Gastroenterology; 2013. Comprehensive genomic meta-analysis identifies intra-tumoural stroma as a predictor of survival in patients with gastric cancer.; p. 1100-11.
7. Perez-Villamil, B.; Romera-Lopez, A.; Hernandez-Prieto, S.; Lopez-Campos, G.; Calles, A.; Lopez-Asenjo, J-A., et al. *BMC Cancer*. Vol. 12. BioMed Central Ltd; 2012. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior.; p. 260
8. Takahashi Y, Ishii G, Taira T, Fujii S, Yanagi S, Hishida T, et al. Fibrous stroma is associated with poorer prognosis in lung squamous cell carcinoma patients. *J Thorac Oncol*. 2011; 6:1460–7. [PubMed: 21849853]
9. Downey, CL.; Simpkins, SA.; White, J.; Holliday, DL.; Jones, JL.; Jordan, LB., et al. *British Journal of Cancer*. Vol. 110. Nature Publishing Group; 2014. The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer.; p. 1744-7.
10. Dekker, TJA.; van de Velde, CJH.; van Pelt, GW.; Kroep, JR.; Julien, J-P.; Smit, VTHBM., et al. *Breast Cancer Res Treat*. Vol. 139. Springer; US: 2013. Prognostic significance of the tumor-stroma ratio: validation study in node-negative premenopausal breast cancer patients from the EORTC perioperative chemotherapy (POP) trial (10854).; p. 371-9.
11. Ahn, S.; Cho, J.; Sung, J.; Lee, JE.; Nam, SJ.; Kim, K-M., et al. *Tumour Biol*. Vol. 33. Springer; Netherlands: 2012. The prognostic significance of tumor-associated stroma in invasive breast carcinoma.; p. 1573-80.
12. Moorman AM, Vink R, Heijmans HJ, van der Palen J, Kouwenhoven EA. The prognostic value of tumour-stroma ratio in triple-negative breast cancer. *Eur J Surg Oncol*. 2012; 38:307–13. [PubMed: 22264965]
13. de Kruijf, EM.; van Nes, JGH.; van de Velde, CJH.; Putter, H.; Smit, VTHBM.; Liefers, GJ., et al. *Breast Cancer Res Treat*. Vol. 125. Springer; US: 2011. Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients.; p. 687-96.
14. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61–70. [PubMed: 23000897]
15. Akbani, R.; Ng, PKS.; Werner, HMJ.; Shahmoradgoli, M.; Zhang, F.; Ju, Z., et al. *Nat Commun*. Vol. 5. Nature Publishing Group; 2014. A pan-cancer proteomic perspective on The Cancer Genome Atlas.; p. 3887
16. Sikora, MJ.; Cooper, KL.; Bahreini, A.; Luthra, S.; Wang, G.; Chandran, UR., et al. *Cancer Res*. Vol. 74. American Association for Cancer Research; 2014. Invasive lobular carcinoma cell lines are characterized by unique estrogen-mediated gene expression patterns and altered tamoxifen response.; p. 1463-74.
17. Huang W, Hennrick K, Drew S. A colorful future of quantitative pathology: validation of Vectra technology using chromogenic multiplexed immunohistochemistry and prostate tissue microarrays. *Hum Pathol*. 2013; 44:29–38. [PubMed: 22944297]
18. Sun, X.; Sandhu, R.; Figueroa, JD.; Gierach, GL.; Sherman, ME.; Troester, MA. *Cancer Epidemiol Biomarkers Prev*. Vol. 23. American Association for Cancer Research; 2014. Benign breast tissue composition in breast cancer patients: association with risk factors, clinical variables, and gene expression.; p. 2810-8.
19. Yamashita, M.; Ogawa, T.; Zhang, X.; Hanamura, N.; Kashikura, Y.; Takamura, M., et al. *Breast Cancer*. Vol. 19. Springer; Japan: 2010. Role of stromal myofibroblasts in invasive breast cancer: stromal expression of alpha-smooth muscle actin correlates with worse clinical outcome.; p. 170-6.
20. Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-García, W., et al. *Nat Commun*. Vol. 4. Nature Publishing Group; 2013. Inferring tumour purity and stromal and immune cell admixture from expression data.; p. 2612
21. Farmer, P.; Bonnefoi, H.; Anderle, P.; Cameron, D.; Wirapati, P.; Wirapati, P., et al. *Nat Med*. Vol. 15. Nature Publishing Group; 2009. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer.; p. 68-74.

22. Byers, LA.; Diao, L.; Wang, J.; Saintigny, P.; Girard, L.; Peyton, M., et al. *Clin Cancer Res.* Vol. 19. American Association for Cancer Research; 2013. An epithelial mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance.; p. 279-90.
23. Hennessy BT, Gonzalez-Angulo AM, Stenke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee J-S, et al. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res.* 2009; 69:4116–24. [PubMed: 19435916]
24. Ono, M.; Tsuda, H.; Shimizu, C.; Yamamoto, S.; Shibata, T.; Yamamoto, H., et al. *Breast Cancer Res Treat.* Vol. 132. Springer; US: 2012. Tumor-infiltrating lymphocytes are correlated with response to neoadjuvant chemotherapy in triple-negative breast cancer.; p. 793-805.
25. Sonntag J, Bender C, Soons Z, Heyde von der S. Reverse phase protein array based tumor profiling identifies a biomarker signature for risk classification of hormone receptor-positive breast cancer. *Transl Proteom.* 2014; 2:52–9.
26. Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. *Trends Mol Med.* 2013; 19:410–7. [PubMed: 23639582]
27. Prat, A.; Adamo, B.; Cheang, MCU.; Anders, CK.; Carey, LA.; Perou, CM. *Oncologist.* Vol. 18. AlphaMed Press; 2013. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer.; p. 123-33.

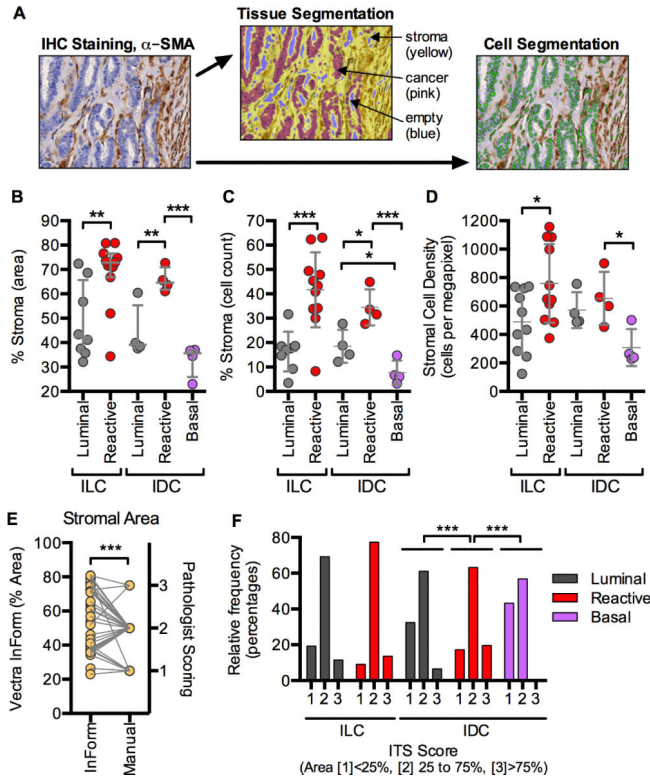


### Translational Relevance

Improved biomarker tests are required to minimize overdiagnosis and overtreatment of breast cancers. A number of pathological criteria have been established to differentiate indolent or aggressive behavior, such as Nottingham grade of cancer cells. However, the effects of the tumor microenvironment on patient outcomes have not been integrated into pathologic criteria. In the current study, the Reactive subtype of breast cancer, identified by reverse phase protein arrays, was demonstrated to indicate a favorable outcome. The lowest risk tumors, typically ER-positive breast cancer tumors, were more likely to have high expression of stromal proteins, high intra-tumoral stromal volume, and increased cell density within the stromal compartment. ER-positive/HER2-negative disease with a Reactive protein signature had a high intra-tumoral stromal content and an excellent prognosis.

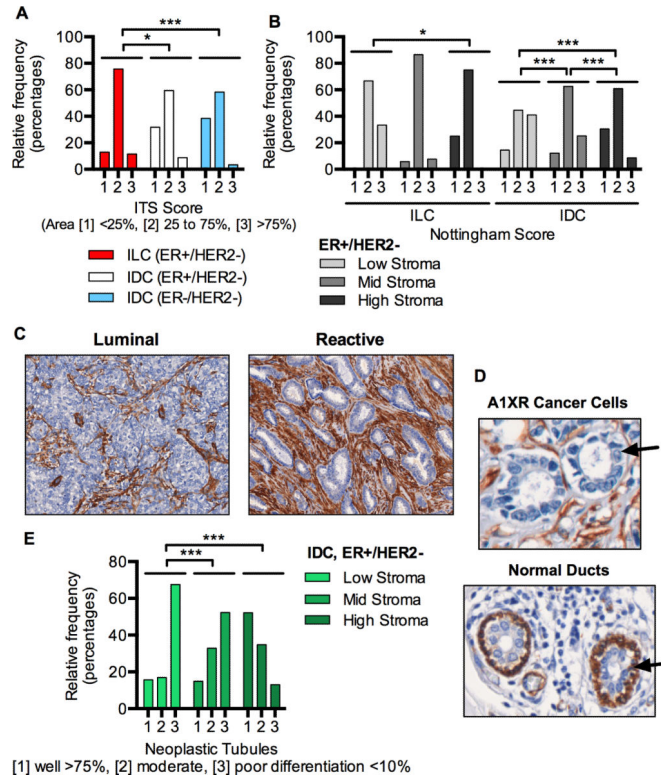


**Figure 1.** Subtyping of breast cancers by RPPA. **A**, Heatmap by unsupervised consensus clustering of protein lysates analyzed for protein expression by RPPA for TCGA breast cancers ( $n = 771$  core cases). Clinical biomarkers: positive (red) or negative (green). Histopathology: IDC (red), ILC (green), or other (grey). mRNA subtyping by PAM50: Luminal A (black), Luminal B (blue), HER2-enriched (orange), Normal-like (green), and Basal-like (purple). *CDH1* mutation: Yes (black), No (grey). Unknowns (white). The approximate locations of clinical biomarkers and other proteins are marked. **B**, Selected differentially-expressed proteins between cluster II (Reactive) and the other clusters. Boxplots are drawn with the Tukey method. ANOVA and Holm-Sidak's multiple comparison tests were performed. Only statistically significant comparisons to cluster II are reported. \*\*\*,  $P < 0.001$ . **C**, Kaplan-Meier survival curves for RPPA clusters in (A).



**Figure 2.**

Reactive breast cancers had the highest stromal content independently of ILC/IDC classification. **A**,  $\alpha$ -SMA IHC staining with tissue and cell segmentation maps for an example case, A1H3. Tissues were segmented into cancer (red), stroma (yellow), empty (blue) using InForm 2.0 (Vectra). Images were selected to include those representing the core invasive tumor, excluding images with edge artifacts and structures unrelated to the tumor including normal ducts and DCIS. Percent stroma determined by area (**B**) and by cell counting (**C**). **D**, Stromal cell density was calculated from the absolute stromal area and cell counts within the compartment. ILC cases were selected from all available sections with clinical data and matching RPPA classification ( $n = 21$ ). ILC includes ILC and ILC mixed cases. Representative IDC cases for RPPA subtypes for (**B**), (**C**), and (**D**) ( $n = 4$  ea; Luminal, Reactive, and Basal) were selected from available cases based on silhouette analysis. **E**, Paired correlation of InForm stromal area from (**B**) and manual pathologist scoring of stromal area. **F**, Stromal content of cancers separated by histopathology and RPPA subtype as determined by manual scoring ( $n = 70$  for ILC,  $n = 331$  for IDC). Symbols and error bars represent the average  $\pm$  one standard deviation. Statistical analyses (paired Student's t-test, unpaired Student's t-test, paired Spearman correlation, or ANOVA as appropriate) were performed within histopathology groups. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$



**Figure 3.** Stromal content association with histopathology, Nottingham grade, and differentiation. **A**, Intra-tumoral stromal (ITS) content of TCGA breast cancers separated by histopathology ( $n = 70$  for ILC,  $n = 352$  for IDC). ILC includes ILC and ILC mixed cases. **B**, Nottingham grade of ER-pos/HER2-neg cases separated by histopathology and stromal content ( $n = 61$  for ILC,  $n = 238$  for IDC). **C**,  $\alpha$ -SMA staining of representative luminal and stromal breast cancers (TCGA cases A1J1 and A1H3) as determined by silhouette analysis. **D**, Cancer cells (top, black arrow marks cancer epithelium) and normal breast ducts (bottom, black arrow marks myoepithelium) from TCGA case A1XR stained with  $\alpha$ -SMA. **E**, Association of ITS score with tubule formation score for TCGA breast cancers ( $n = 261$ , IDC, ER-pos/HER2-neg). ANOVA and Holm-Sidak's multiple comparison tests were performed for (A), (B), and (E) assuming the pathology score was a continuous variable. \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$

**Table 1**

Association of clinical characteristics to the RPPA Reactive subtype for ER-positive/HER2-negative and all breast cancers, TCGA core cohort.

Characteristic	ER-pos/HER2-neg			All		
	No. Reactive	%	$\chi^2 P$	No. Reactive	%	$\chi^2 P$
No. of Patients	125	29.9	–	210	27.1	–
Age			0.75			0.831
< 50 years	35	31.5		58	26.4	
50 years	90	29.3		152	27.4	
Tumor size at diagnosis (TNM)			0.16			0.522
T0/T1/T2	97	28.3		163	26.0	
T3/T4	28	37.3		38	29.2	
Lymph Node			1.0			0.441
Negative	49	30.8		90	29.4	
Positive	63	31.3		88	26.2	
Histopathology			<0.001			<0.001
IDC	62	21.5		111	20.1	
ILC/Mixed	57	57.0		77	55.0	
Nottingham Grade *			0.009			<0.001
1	14	31.8		15	28.8	
2	68	33.7		96	31.2	
3	14	17.3		32	14.3	
PAM50 Subtype **			<0.001			<0.001
Luminal A	104	36.2		139	36.7	
Luminal B	12	10.9		18	10.5	
HER2-enriched	0	0.0		4	7.5	
Normal	8	88.9		18	81.8	
Basal	0	0.0		16	12.7	
TP53 mutation			0.028			<0.001
WT	104	32.7		150	30.5	
Mut	12	18.2		29	14.6	
CDHI mutation			<0.001			<0.001
WT	87	26.3		139	22.6	
Mut	29	54.7		40	54.1	

\* For Nottingham grading, grades 1 and 2 pooled were compared to grade 3 cases.

\*\* For PAM50 intrinsic subtyping, Luminal A subtype was compared to other subtypes pooled (Luminal B, HER2-enriched, and Basal).

Association of clinical characteristics and other subtyping to patient outcomes for ER-positive/HER2-negative and all breast cancers, TCGA core cohort.

**Table 2**

Characteristic	ER-pos/HER2-neg					All					
	No.	Hazard Ratio	95% CI for HR			No.	Hazard Ratio	95% CI for HR			
			Low	High	P			Low	High	P	
No. of Patients	418					774					
Overall survival											
Event	34					85					
No event	384					678					
Age				0.051							0.130
< 50 years	111	1			220	1					
50 years	307	3.11	0.94	10.35	554	1.54	0.88	2.71			
Tumor size at diagnosis (TNM)				0.234							0.028
T0/T1/T2	343	1			626	1					
T3/T4	75	1.68	0.71	3.98	130	1.82	1.06	3.14			
Lymph Node				0.684							0.054
Negative	159	1			306	1					
Positive	201	1.18	0.53	2.66	336	1.71	0.99	2.96			
Histopathology				0.111							0.032
IDC	289	3.1	0.72	13.10	550	2.9	1.05	7.96			
ILC/Mixed	100	1			140	1					
Nottingham Grade				0.547							0.038
[1,2]	246	1			306	1					
[3]	81	0.69	0.20	2.33	224	1.74	1.02	2.94			
PAM50 Intrinsic Subtype				0.292							0.009
Luminal A	287	0.66	0.30	1.44	379	0.52	0.31	0.86			
Luminal B/HER2-enriched/Basal	120	1			340	1					
TP53 mutation				0.060							0.043
WT	318	1			480	1					
Mut	66	2.21	0.95	5.14	196	1.68	1.01	2.79			
CDH1 mutation				0.250							0.196



Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Characteristic	ER-pos/HER2-neg					All				
	No.	Hazard Ratio	95% CI for HR		P	No.	Hazard Ratio	95% CI for HR		P
			Low	High				Low	High	
WT	331	1				607	1			
Mut	53	0.327	0.04	2.42		69	0.47	0.15	2.79	
RPPA Subtype					0.049					0.028
Reactive	125	0.36	0.12	1.04		210	0.48	0.24	0.94	
Other	293	1				564	1			