

**PHS PUBLIC ACCESS**

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2016 February 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2015 February ; 24(2): 406–414. doi: 10.1158/1055-9965.EPI-14-0934.**Tumor Intrinsic Subtype is Reflected in Cancer-Adjacent Tissue****Patricia Casbas-Hernandez¹, Xuezheng Sun², Erick Roman-Perez², Monica D'Arcy², Rupninder Sandhu³, Asahi Hishida³, Kirk K. McNaughton⁴, Xiaohong R. Yang⁵, Liza Makowski^{3,6}, Mark E. Sherman⁵, Jonine D. Figueroa⁵, and Melissa A. Troester^{1,2,3}**

Patricia Casbas-Hernandez: casbas@email.unc.edu; Xuezheng Sun: amysun@email.unc.edu; Erick Roman-Perez: erick.roman@med.navy.mil; Monica D'Arcy: mdarcy@email.unc.edu; Rupninder Sandhu: sandhu@email.unc.edu; Asahi Hishida: hishida@email.unc.edu; Kirk K. McNaughton: Kirk_McNaughton@med.unc.edu; Xiaohong R. Yang: royang@mail.nih.gov; Liza Makowski: liza.makowski@unc.edu; Mark E. Sherman: Sherman@exchange.nih.gov; Jonine D. Figueroa: figueroaj@mail.nih.gov; Melissa A. Troester: Troester@email.unc.edu

¹Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599²Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599³Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599⁴Department of Physiology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599⁵Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA⁶Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599**Abstract****Introduction**—Overall survival of early-stage breast cancer (BC) patients is similar for those who undergo breast conserving therapy (BCT) and mastectomy, however, 10-15% of women undergoing BCT suffer ipsilateral breast tumor recurrence. The risk of recurrence may vary with BC subtype. Understanding the gene expression of the cancer-adjacent tissue and the stromal response to specific tumor subtypes is important for developing clinical strategies to reduce recurrence risk.**Methods**—We utilized two independent data sets to study gene expression data in cancer-adjacent tissue from invasive BC patients. Complementary *in vitro* cocultures were used to study cell-cell communication between fibroblasts and specific BC subtypes.**Results**—Our results suggest that intrinsic tumor subtypes are reflected in histologically-normal cancer-adjacent tissue. Gene expression of cancer-adjacent tissues shows that triple negative (Claudin-low or Basal-like) tumors exhibit increased expression of genes involved in inflammation and immune response. While such changes could reflect distinct immuneCorrespondence to: Melissa A. Troester, Troester@email.unc.edu.**Conflict of interest:** The authors declare no conflict of interest.

populations present in the microenvironment, altered immune response gene expression was also observed in cocultures in the absence of immune cell infiltrates, emphasizing that these inflammatory mediators are secreted by breast-specific cells. In addition, while triple negative BCs are associated with upregulated immune response genes, Luminal breast cancers are more commonly associated with estrogen-response pathways in adjacent tissues.

Conclusions—Specific characteristics of BCs are reflected in the surrounding histologically-normal tissue. This commonality between tumor and cancer-adjacent tissue may underlie second primaries and local recurrences.

Impact—Biomarkers derived from cancer-adjacent tissue may be helpful in defining personalized surgical strategies or in predicting recurrence risk.

Introduction

Breast conservation therapy (BCT) with lumpectomy and radiotherapy and mastectomy are equally effective in treating early stage breast cancer. However, approximately 10-15% of women undergoing BCT suffer ipsilateral breast tumor recurrence [1-3], often with metastasis [4, 5]. Younger age has been associated with higher recurrence [6], but tumor characteristics may also be responsible because aggressive breast cancers tend to be diagnosed in younger women [7] and have higher local recurrence rates [8, 9].

Breast stromal microenvironments (including fibroblasts, endothelial cells, and immune cells), change during carcinogenesis. Cancer-associated fibroblasts may play a critical role in maintaining chronic inflammation around breast cancers [10], and may also have regulatory effects independent of immune cells [11-13]. Stromal microenvironments also vary by breast cancer subtype, and may influence progression [14] [15, 16]. Recent studies have examined benign, cancer-adjacent tissue and found substantial interindividual variation. These studies show two distinct subtypes of cancer-adjacent tissues with distinct survival patterns [17], and show that “molecular histology” of epithelium in cancer-adjacent tissues surrounding Estrogen Receptor (ER)-negative tumors differ from those of ER-positive cancers [18]; ER-positive tumors are associated with high expression of ER mRNA in cancer-adjacent tissue [19]. Thus, understanding the microenvironment surrounding breast cancer subtypes is important for recurrence and in targeting surgical strategies.

We hypothesized that genomic features of histologically-normal, cancer-adjacent tissue differ by intrinsic subtype. Based on previous findings from cell culture and mouse models showing upregulation of key chemokines and growth factors in fibroblast interactions with basal-like breast cancers [14, 20, 21], it is important to characterize the microenvironment response to basal-like breast cancer in human tissue. We also sought to validate previous reports of differences in estrogen responsiveness of ER-positive tumor-adjacent tissue. We therefore investigated gene expression profiles of cancer-adjacent tissue using data from two independent sources: the National Cancer Institute's Polish Breast Cancer Study and The Cancer Genome Atlas (TCGA) Project. We then used insights from these *in vivo* studies to further interrogate subtype-specific gene expression *in vitro*.

Materials and Methods

Polish Women's Breast Cancer Study (PWBCS), the TCGA Study and the Normal Breast Study (UNC)

This study included 139 women from the PWBCS with snap-frozen extratumoral breast and tumor tissues (Supplemental Table 1). The PWBCS is a population-based case-control study conducted in Poland (Warsaw and Łódź) during 2000-2003 [22]. PWBCS cases were women aged 20-74 years with pathologically-confirmed *in situ* or invasive breast carcinoma. Tissues from invasive tumors and non-neoplastic cancer-adjacent breast tissue were collected at the time of breast surgery. Histologically-normal, cancer-adjacent tissues were <2 cm from the tumor margin. Based on *in vitro* evidence of their distinctive microenvironments, Basal-like and Luminal tumors were oversampled for these analyses. Patient data was collected from medical records and in-person interviews as described previously. All participants provided written informed consent under a protocol approved by the National Cancer Institute and Polish institutional review boards.

An additional 60 snap-frozen cancer-adjacent samples collected and analyzed by the TCGA were used as a validation data set (Supplemental Table 1). These samples were all histologically-normal, cancer-adjacent (<2cm from tumor margin) to invasive breast carcinoma and tumor subtype was classified for these samples and reported previously [23].

A subset of 36 cancer-adjacent samples from 'The Normal Breast Study' (NBS) were used for this study to evaluate distance from tumor margin. NBS is a hospital-based cross-sectional study conducted in UNC-Hospitals (Chapel Hill, NC, USA) from 2009-2014. All patients had a newly diagnosed invasive breast carcinoma. Fresh tissues were collected at the time of breast surgery and snap frozen in liquid nitrogen. Tumor adjacent breast tissues used in this study were classified as peritumoral (<2 cm from the tumor margin) and remote (>2 cm from the tumor margin). Information on clinicopathological, demographic, and anthropometric factors was collected from medical records and in-person interviews. Detailed data on tumor subtype was unavailable for these patients so they could not be used for primary analyses of subtype-specific microenvironments. All of the participants provided written informed consent under a protocol approved by the IRB.

Tumor expression analysis: Molecular classification of the tumor using PAM50

Tumor samples from PWBCS and TCGA were used to determine molecular subtype of the cancer. RNA was isolated using previously published methods. For PWBCS data, Illumina Ref-8 Beadchip Version2 microarray platform was used and data normalization was performed using Lumi in R. For TCGA data, custom Agilent arrays or RNA sequencing were performed as described in [23]. To classify tumors, genes were median-centered and samples were standardized to zero mean and unit variance. The PAM50 predictor was performed [24] to categorize the tumors into five subtypes (Luminal A, Luminal B, Her2-enriched, Basal-like, and normal-like). The Claudin-low predictor was applied as in [25].

Cancer-adjacent expression analysis

For PWBCS, two-color 4X44K Agilent whole genome arrays were performed on a frozen section of cancer-adjacent tissue, with sections on either side used for imaging of cellular composition. Tissue for microarrays was homogenized using a MagnaLyser homogenizer (Roche), and RNA was isolated and quality was checked as described in Troester et al. [15]. Microarrays were performed as previously described [26]. Briefly, Cy3-labeled reference was produced from total RNA from Stratagene Universal Human Reference (spiked 1:1,000 with MCF-7 RNA and 1:1,000 with ME16C RNA to increase expression of breast cancer genes) following amplification with Agilent low RNA input amplification kit. Patient samples were labeled with Cy5. Data were Lowess normalized, and probes with a signal <10 dpi in either channel were excluded as missing. Probes with more than 20% missing data across samples were excluded. In data preprocessing, we (1) eliminated probes without corresponding ENTREZ ID, (2) collapsed duplicate probes by averaging, (3) imputed missing data using k-nearest neighbors (KNN) method with k=10, and (4) median-centered genes. Microarray data are publicly available through the Gene Expression Omnibus (GEO) (GSE49175 [26], GSE50939). TCGA data and methods are available at the TCGA Data Portal (<https://cga-data.nci.nih.gov>).

Supervised analysis of cancer-adjacent tissue

Using the PWBCS as a training set, four-class significance analysis of microarrays (SAM) [27] was used to identify differentially expressed genes associated with breast cancer subtypes (LuminalA, LuminalB, Her2 and Basal-like plus Claudin-low) [27]. Significance was defined as False Discovery Rate (FDR) 0.1%. Tumors classified as “normal-like” may result from extensive normal or stromal content in the tumor [28], so we excluded normal-like tumors. The genes identified as differentially expressed in these four groups from the SAM analysis are henceforth referred to as the ‘*In vivo triple negative microenvironment signature*’ (Figure 1; Full list of genes can be found in Supplemental Table 2). Because there was a common pattern for both triple negative tumor subtypes vs. other subtypes, we collapsed tumor subtypes to conduct a 2-class comparison (Basal-Like/Claudin low vs. HER2/Luminal). This 2-class gene list was used only for gene ontology; the more parsimonious (fewer genes) 4-class list was used for all subsequent classification. Gene ontology analysis was done using Ingenuity Pathway Analysis with Benjamini–Hochberg multiple testing correction to identify significant functions and pathways (P -values < 0.05). Pathways and functions with less than 2 genes were excluded (Supplemental tables 3 and 4).

To evaluate the association of gene expression for each patient with the defined biological signature, Pearson correlation coefficients were obtained as described in [17, 26, 29]. Briefly, for a given gene signature (i.e., EReS [30] and the ‘*In vivo triple negative microenvironment signature*’), ‘1’ was assigned to up-regulated and ‘-1’ to downregulated genes, Pearson correlation coefficients were calculated by comparing this standard vector to the measured, median-centered gene expression level for each patient. Patients were classified as positive if the Pearson correlation coefficient was ≥ 0 , and negative if the coefficient was <0. These classes were further evaluated for their association with other tumor characteristics. The EReS signature was analyzed in both data sets and the ‘*In vivo triple negative microenvironment signature*’ was identified in the PWBCS and tested in the

TCGA dataset. Because all analyses were conducted with median-centered data, expression is relative to other subtypes. That is, high expression of the Triple Negative signature in tissue adjacent to Basal-like and Claudin-low tumors implies lower expression of these inflammatory genes among tissue adjacent to Luminal tumors. Conversely, high expression of the EReS in tissue adjacent to Luminal tumors implies relatively lower estrogen responsiveness among tissue adjacent to triple negative tumors.

Composition analysis of cancer-adjacent tissues from PWBCS

Frozen sections were obtained from the same frozen block used for the microarray analysis and were taken immediately adjacent to the piece used for RNA extraction. 127 samples from the PWBCS and 47 samples from TCGA had hemotoxylin and eosin (H&E) stained sections of sufficient quality to be analyzed for tissue composition. Composition analysis was performed as explained in [26]. Briefly, 20 μm (PWBCS) or 5 μm (TCGA) frozen sections were H&E stained. A training set of slides was scanned using Aperio ScanScope CS V11.0.2.725 and images were manually annotated for composition of adipose, epithelium and non-fatty stroma. This training set was used to create a Genie (Aperio Technologies, Vista, CA, USA) algorithm with high accuracy in segmenting adipose tissue, epithelium, non-fatty stroma, and glass on each slide. Agreement between manual and digital assessment exceeded 98%, and thus the data from Genie digital segmentation was used in analyses. Composition data from both data sets (mean values) were used to determine a cut-point for dichotomized analyses of composition; an epithelium cut-point of 10% (Mean values PWBCS=9.8% and TCGA=10.2%) and a stroma cut-point of 20% (Mean values PWBCS=26.8% and TCGA= 14.4%) were selected.

Cell lines and Coculture conditions

Cell lines were purchased and maintained as in [14]; MCF7, SKBR3, MDA-MB-231, HCC1937 and MDA-MB-468 were purchased from ATCC and passaged for less than 6 months prior to being used for experiments, Sum159 were purchased from Asterand and passaged for less than 6 months prior to being used for experiments, authentication was provided at time of purchase. ZR75, T47D, Sum149, ME16C and Sum102 were described in Troester *et al.* [31]. *Direct cocultures*, wherein cells are seeded in a single well and in physical contact (not separated by transwell membranes), were performed as previously described [14]. Briefly cancer cells lines (MCF7, ZR75, T47D, SKBR3, MDA-MB-231, SUM159, SUM149, HCC1937, ME16C, SUM102 and MDA-MB-468) and immortalized reduction mammary fibroblasts (RMFs, [32]) were plated on plastic in direct contact, potentially interacting both through secreted factors and through cell-cell contact. RMFs were tested for viability in all cancer cell media, and direct cocultures were maintained in the appropriate cancer cell media (e.g., MCF-7 in RPMI). The following RMF:cancer cell ratios were plated for most direct cocultures: 0:1, 1:4, 1:2, 1:1, 2:1, 1:0. Cocultures and monocultures (for comparison) were maintained for 48 hours prior to RNA isolation.

RNA and expression microarrays: cell lines

Monocultures and cocultures were harvested by scraping in RNA lysis buffer. Total RNA was isolated using the RNeasy mini kit (Qiagen) as previously described in (Camp et al

[14]). Microarrays were performed according to Agilent protocol using 2-color Agilent 4×44K (Agilent G4112F) human arrays and 244K (Agilent G4502A) custom human arrays. Only probes present on both platforms were utilized. Samples were run in batches together with appropriate monoculture controls to minimize the effect of batch. Agilent's Quick Amp labeling kit was used to synthesize Cy3-labeled reference from Stratagene Universal Human Reference (as described above) and Cy5-labeled RNA from cocultured or monocultured cell lines (as previously described in [14]). Data are available through the GEO (GSE26411).

Coculture data normalization and analysis

Data from 122 microarrays (representing monocultures and direct cocultures from 12 different cell lines described above) were included. Only genes where >70% of microarrays had signal >10 dpi in both channels were included. Data were Lowess normalized and missing data were imputed using *k*-nearest neighbors' imputation. For the *direct coculture* analyses, we excluded genes that did not have at least 2-fold deviation from the mean in at least 1 sample and the method of Buess *et al.*[33, 34] was used to normalize cocultures to appropriate monocultures as described in [14]. Briefly, the Buess method is an established expression deconvolution approach for direct cocultures of 2 different cell lines that estimates the percent of fibroblasts and cancer cells in each coculture, and normalizes the data for composition differences prior to estimating the effect of epithelial-stromal interaction on gene expression. The Buess interaction coefficient “I” was calculated as the ratio of observed to expected gene expression for each gene and an “I-matrix” representing the epithelial-stromal interaction coefficients for each gene in each coculture was generated. This I-matrix was analyzed using multiclass Significance Analysis of Microarrays (SAM [27]) in R.1.14, comparing Basal-like and Claudin-low cocultures (i.e. direct cocultures of MBA-MB-231, SUM159, SUM149, HCC1937, SUM102 and MDA-MB-468 with RMFs) to Her2-enriched and Luminal cocultures (direct cocultures of MCF7, ZR75, T47D, SKBR3, ME16C with RMFs). Heatmaps were generated using Cluster 3.0 and Java treeview was used to visualize data.

Evaluating correlation between breast cancer subtype and *in vitro* triple negative gene signature

The ‘*in vitro* triple negative signature’ (Supplemental table 2) was defined by all genes significantly upregulated in basal-like and claudin-low cocultures after performing a multiclass SAM analysis. Because, all of the genes defining this ‘*in vitro* triple-negative signature’ were upregulated, the average expression levels across all genes was used to score tumors and normal tissue for the triple negative signature. Expression values were median-centered by gene before summing. Boxplots were used to compare the triple negative score across intrinsic subtypes. We further evaluated the difference in mean expression by subtype using Analysis of Variance (ANOVA).

Statistical analysis

R version 1.14 was used to generate box plots, evaluate signature correlations, and to perform chi-square significance tests. SAS 9.2 (32) was used for logistic regression to estimate odds ratios for expression of estrogen response and the triple negative stromal

response signatures by subtype. Fisher's Exact test was used to test associations with clinical features and for subtype-stratified tissue composition analysis.

Results

Tumor Intrinsic Subtype is reflected in Cancer-Adjacent Tissue

We used samples from the Polish Women's Breast Cancer Study (PWBCS, population characteristics in Supplemental Table 1) to identify subtype-associated changes in cancer-adjacent tissue. Triple negative cancer-adjacent tissues had a unique stromal response, with Figure 1 showing a heatmap of 126 genes (Supplemental Table 2) whose expression differed significantly between tissues adjacent to Basal-like, Claudin-low, Luminal B, and Luminal A tumors. Gene ontology analysis (using Ingenuity Pathway Analysis) revealed that genes upregulated in tissues adjacent to Triple Negative tumors are involved in functions and pathways such as activation of leukocytes, proliferation of mononuclear leukocytes, cell movement of leukocytes, interferon signaling, Hepatic Fibrosis, T-helper cell differentiation or antigen presentation pathway (Full list in Supplemental Tables 3 and 4).

Our results suggest that the cancer-adjacent tissue shares biology of the tumors themselves. Four genes (NAT1, FOXA1, MLPH, ESR1) used in the PAM50 subtyping [24] have differential expression in the tissue adjacent to Luminal breast cancers. Having observed high expression of estrogen receptor 1 (ESR1) adjacent to Luminal breast cancers, and in light of previous reports suggesting similarities between ER positive tumors and their adjacent tissue [19], we utilized a published estrogen response signature (EReS) [30] to characterize the estrogen response of each cancer-adjacent tissue (positive or negative for the EReS) in both populations. Table 1 shows that in both populations there was a significant association between the expression of this signature and breast cancer subtype, with the majority of Luminal A cancer-adjacent tissue being positive for EReS (PWBCS: 62.26% and TCGA: 68.00%) and the vast majority of the more aggressive Claudin-low cancer-adjacent tissue being negative (PWBCS: 92.31% and TCGA: 100.00%). Moreover, there was a significant association between expression of the *in vivo* triple negative signature by subtype (*p-value*: 0.0033 in PWBCS, training set and *p-value*: 0.0005 in TCGA, test set, respectively). Claudin-low cancer-adjacent tissues had the highest percentage of '*in vivo* triple negative signature'-positive tumors (84.65% in PWBCS and 100% in TCGA) and Luminal A cancer-adjacent tissues had the lowest percentage of positive tumors (only 34.00% in PWBCS and 28.00% in TCGA) (Table1).

Cancer-adjacent expression is not associated with tissue composition

One source of variation in cancer-adjacent tissue is the heterogeneous composition, with some patients having more or less stroma and epithelium than others. To evaluate the role of tissue composition in expression of the *in vivo* triple negative microenvironment signature, we quantified the proportion of epithelium, non-fatty stroma and fatty/adipose tissue in each sample. Histological sections adjacent to the portion of tissue used for RNA extraction were used for H&E staining and analysis. Table 2 dichotomizes the samples according to composition. There were no statistically significant associations between tumor subtype (from which the adjacent tissue came) and epithelial content (Fisher's Exact *p-value*=0.213

and p -value=0.177) or non-fatty stromal content (Fisher's Exact p -value=0.235 and p -value=0.104) in either dataset of cancer-adjacent tissue, underscoring that the triple negative gene expression profile is not due to differences in composition by subtype. However the contrary is true for the EReS signature, it was significantly associated with non-fatty stromal content (Chi square p -value<0.001 in PWBCS and p -value = 0.049 in TCGA) in both datasets, suggesting a role for the stroma in estrogen response of normal breast tissue.

Given that tumor margins may vary from tumor to tumor, we addressed the EReS signature in an additional dataset from UNC-hospitals to evaluate whether distance from margin could represent an unmeasured confounder of these analyses. We obtained 32 ER-positive, cancer-adjacent samples, 16 each from <2cm and >2 cm from the tumor. The majority of these samples were positive for the EReS signature (10 of 16 were positive at both distances). This suggests that the correlation is not strongly dependent upon distance to tumor.

Cancer-adjacent biology can be recapitulated in vitro

We next identified a triple negative signature in coculture (*in vitro*) and evaluated whether this *in vitro* signature also accurately identifies triple negative-adjacent samples. Twelve breast cancer subtype models were cocultured with immortalized reduction mammary fibroblasts (RMFs) and we identified a unique set of genes upregulated in triple negative cocultures (Figure 2A, grey bar 'In vitro triple negative signature', full list of genes can be found in Supplemental Table 2). To address whether this *in vitro* gene signature is associated with tissue adjacent to triple negative tumors two approaches were used: (1) comparison of pathways and biological functions obtained *in vitro* vs. in tissue and (2) evaluation of the *in vitro* signatures in breast tumors and the cancer adjacent tissue.

Using the first approach, genes identified through the *in vitro* cocultures (Supplemental Tables 5 and 6) and *in vivo* (Supplemental Tables 3 and 4) were in similar pathways. Statistically significant biological functions and pathways in common were activation of cells, proliferation of mononuclear leukocytes, cell movement of leukocytes, inflammatory response, hepatic fibrosis, role of cytokines in mediating communication between immune cells and IL-6 signaling. Using the second approach, Basal-like and Claudin-low tumors had high expression of genes identified in cocultures of triple negative cancer cell lines, both within the tumor (Figure 2B, one-way ANOVA by subtype, $p=8.44e-15$) and in the cancer-adjacent tissue (of the PWBCS) (Figure 2C). The association with cancer-adjacent tissue was weaker and not significant (p -value=0.196), however, the expression from cancer-adjacent tissues qualitatively mirrors the expression patterns from tumors.

Discussion

Studies of breast cancer microenvironment by subtype [19, 35] have important implications for local recurrence. Locoregional recurrence may be higher among Basal-like breast cancers [36], and we hypothesized that these cancers may induce a permissive microenvironment for local recurrence. The results presented in this manuscript suggest that cancer-adjacent tissue of Basal-like and Claudin-low breast cancers differs substantially from that of Luminal cancers and that these differences are strongly dependent upon

fibroblast interactions and/or stromal composition. That is, the microenvironment may be primed for inflammation [37] or estrogen response by the extratumoral stroma.

Previous studies have isolated intratumoral and extratumoral stroma and evaluated gene expression in association with outcome. For example, Finak *et al.* reported that stroma adjacent to basal-like breast cancers had high levels of immune response genes and that these genes predicted progression [16]. However, microdissection cannot perfectly exclude immune infiltrates, so it was difficult to identify the cell type responsible for the up-regulation of immune mediators. Our use of cocultures allowed us to identify fibroblasts as key contributors to cytokine/chemokine expression. Even in the absence of immune cells, fibroblasts and epithelial cells produce molecules that affect immune cell recruitment and activation and that directly regulate epithelial cell differentiation (e.g. IL-6 alters epithelial cell phenotypes [38]). Thus, while previous studies established that Triple Negative BCs are associated with a pro-inflammatory milieu [25, 37, 39, 40], the current findings suggest that this reaction may initiate in epithelial-fibroblast interactions and occurs in both tumors and surrounding histologically-normal tissue.

Our results also addressed unique features of Luminal microenvironments *in vivo*. Cancer-adjacent gene expression for Luminal tumor subtypes is markedly different from that of Basal-like and Claudin-low tumors, with Luminal-adjacent tissues expressing high levels of Luminal breast tumor markers [24]. In other words, the cancer-adjacent tissue around the tumor reflects or may even predict the biology of the tumor that arises. If ER-positive tumors are more likely to occur in pervasively estrogen responsive benign tissues [41, 42], extratumoral signatures could be candidate biomarkers for predicting subtype-specific risk. Others have hypothesized that host factors or widespread field effects cause first and second primaries to have similar phenotypes, with most second primaries having the same phenotype as the first primary [43, 44]. Thus, the idea of an estrogen responsive “field effect” is supported by different types of human data. The viability of using estrogen response in normal tissue to predict risk or recurrence risk depends upon whether these signatures occur early or late in carcinogenesis. Because our data was collected at a single time point after tumor onset, the genomic differences we observed could represent (1) patterns of predisposition that pre-exist tumor formation, or (2) a reaction to the tumor. We hypothesize that the triple negative signatures are a response to tumor, while the estrogen response signature reflects susceptibility. Future longitudinal studies are needed to test these hypotheses in additional populations and should consider other epidemiologic variables (such as BMI, age or estrogen exposure).

Cancer-adjacent tissue is composed of a high percentage of stroma (both fatty and non-fatty) and gene expression analyses of these tissues are enriched for stromal pathways. Because of this, other studies have approached the study of adjacent tissue by microdissecting and studying individual cellular components [16, 45, 46]. However, cellular heterogeneity is important, and inclusion of all cell types enabled us to evaluate how stromal composition relates to pathway expression. Our coculture results confirm that immune response may initiate with stromal-epithelial interactions. Our tissue studies show concordance between EReS and stromal content, suggesting a role for stroma in modulating estrogen activity. Previous literature also suggests biological plausibility for this association [47], but the

mechanisms of estrogen action in stroma remain to be fully elucidated. However, regardless of the ultimate mechanisms, both of the signatures evaluated were originally defined in cell lines, underscoring our findings and those of others that show that cell line-derived signatures have accuracy in elucidating *in vivo* biology [14, 33, 48, 49].

Strengths of this analysis include use of two distinct sample sets: one as a training dataset (PWBCS), and the other as an independent validation set (TCGA). In addition, tissue composition was considered as a potential modifier. It is also a strength that we have validated the signatures in a controlled, experimental coculture system. Future work should focus on investigating the role of distance from tumor more thoroughly, preferably in larger study populations.

In conclusion, we found distinct biological characteristics of cancer-adjacent tissue depending upon tumor intrinsic subtype. This commonality between tumor and surrounding tissue may underlie second primaries and provides plausible explanations for local recurrence. These results also suggest that tissue biomarkers derived from cancer-adjacent tissue may help in predicting risk and in defining appropriate, personalized surgical strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: The authors would like to acknowledge and thank Dr. Keith Amos for all his help and insight during the initial stages of this manuscript. The authors gratefully acknowledge the support of Avon Foundation, the University Cancer Research Fund, the National Cancer Institute (R01CA-138255), the NCI/National Institutes of Environmental Health Sciences (NIEHS) Breast Cancer and the Environment Research Program (BCERP) (U01-ES-019472), pilot funds from the Carolina Cancer Center for Nanotechnology Excellence (U54CA119343) and the SPORE Grant (P50-CA-058223). P.C.H. was supported by a Howard Hughes Medical Institute Fellowship. The Polish Breast Cancer Study was supported by the National Cancer Institute Intramural funds. We would also like to acknowledge Montse Garcia-Closas, Louise Brinton and Jolanta Lissowska for contributions to the Polish study design; and Paul Meltzer, Sean Davis and Sarah Anzick for profiling data of tumor tissues. We would also like to acknowledge the UNC Translational Pathology Laboratory for their contribution with scoring and analyzing the histology slides.

Financial support: Xiaohong R. Yang, Mark E. Sherman and Jonine D. Figueroa received NCI intramural funds. Melissa A. Troester received funds from the Avon Foundation, the University Cancer Research Fund, the National Cancer Institute (R01CA-138255), the NCI/National Institutes of Environmental Health Sciences (NIEHS) Breast Cancer and the Environment Research Program (BCERP) (U01-ES-019472), pilot funds from the Carolina Cancer Center for Nanotechnology Excellence (U54CA119343) and the SPORE Grant (P50-CA-058223).

References

1. Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, Aguilar M, Marubini E. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *New England Journal of Medicine*. 2002; 347:1227–1232. [PubMed: 12393819]
2. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, Jeong JH, Wolmark N. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *New England Journal of Medicine*. 2002; 347:1233–1241. [PubMed: 12393820]

3. Silverstein MJ, Lagios MD, Groshen S, Waisman JR, Lewinsky BS, Martino S, Gamagami P, Colburn WJ. The influence of margin width on local control of ductal carcinoma in situ of the breast. *New England Journal of Medicine*. 1999; 340:1455–1461. [PubMed: 10320383]
4. Vargas C, K L, Go N, Krauss D, Chen P, Goldstein N, Martinez A, Vicini FA. Factors associated with local recurrence and cause-specific survival in patients with ductal carcinoma in situ of the breast treated with breast-conserving therapy or mastectomy. *Int J Radiat Oncol Biol Phys*. 2005; 63(5):1514–1521. [PubMed: 16005576]
5. Bijker N, P J, Duchateau L, Julien JP, Fentiman IS, Duval C, Di Palma S, Simony-Lafontaine J, Mascarel I, Van de Vijer MJ. Risk factor for recurrence and metastasis after bct for dcis: Analysis of european organization for research and treatment of cancer trial 10853. *Journal of Clinical Oncology*. 2001; 19:2263–2271. [PubMed: 11304780]
6. Arvold ND, Taghian AG, Niemierko A, Abi Raad RF, Sreedhara M, Nguyen PL, Bellon JR, Wong JS, Smith BL, Harris JR. Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy. *Journal of Clinical Oncology*. 2011; 29:3885–3891. [PubMed: 21900114]
7. Pirone JR, D'Arcy M, Stewart DA, Hines WC, Johnson M, Gould MN, Yaswen P, Jerry DJ, Schneider SS, Troester MA. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. *Cancer Epidemiology Biomarkers & Prevention*. 2012; 21:1735–1744.
8. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. *Journal of Clinical Oncology*. 2010; 28:1684–1691. [PubMed: 20194857]
9. Gabos Z, Thoms J, Ghosh S, Hanson J, DeschÃˆnes J, Sabri S, Abdulkarim B. The association between biological subtype and locoregional recurrence in newly diagnosed breast cancer. *Breast Cancer Research and Treatment*. 2010; 124:187–194. [PubMed: 20814819]
10. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer*. 2006; 6:392–401. [PubMed: 16572188]
11. Albini A, Sporn MB. The tumour microenvironment as a target for chemoprevention. *Nat Rev Cancer*. 2007; 7:139–147. [PubMed: 17218951]
12. De Wever O, Demetter P, Mareel M, Bracke M. Stromal myofibroblasts are drivers of invasive cancer growth. *International Journal of Cancer*. 2008; 123:2229–2238.
13. Bissell MJ, Kenny PA, Radisky DC. Microenvironmental regulators of tissue structure and function also regulate tumor induction and progression: The role of extracellular matrix and its degrading enzymes. *Cold Spring Harbor Symposia on Quantitative Biology*. 2005; 70:343–356.
14. Camp JT, Elloumi F, Roman-Perez E, Rein J, Stewart DA, Harrell JC, Perou CM, Troester MA. Interactions with fibroblasts are distinct in basal-like and luminal breast cancers. *Molecular Cancer Research*. 2011; 9:3–13. [PubMed: 21131600]
15. Troester MA, Lee MH, Carter M, Fan C, Cowan DW, Perez ER, Pirone JR, Perou CM, Jerry DJ, Schneider SS. Activation of host wound responses in breast cancer microenvironment. *Clinical Cancer Research*. 2009; 15:7020–7028. [PubMed: 19887484]
16. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, Chen H, Omeroglu G, Meterissian S, Omeroglu A, Hallett M, Park M. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med*. 2008; 14:518–527. [PubMed: 18438415]
17. Roman-Perez E, Casbas-Hernandez P, Pirone J, Rein J, Carey L, Lubet R, Mani S, Amos K, Troester M. Gene expression in extratumoral microenvironment predicts clinical outcome in breast cancer patients. *Breast Cancer Research*. 2011; 14:R51. [PubMed: 22429463]
18. Yang X, Figueroa J, Falk R, Zhang H, Pfeiffer R, Hewitt S, Lissowska J, Peplonska B, Brinton L, Garcia-Closas M, Sherman M. Analysis of terminal duct lobular unit involution in luminal a and basal breast cancers. *Breast Cancer Research*. 2013; 14:R64. [PubMed: 22513288]
19. Graham K, Ge X, de las Morenas A, Tripathi A, Rosenberg CL. Gene expression profiles of estrogen receptor positive and estrogen receptor negative breast cancers are detectable in histologically normal breast epithelium. *Clinical Cancer Research*. 2011; 17:236–246. [PubMed: 21059815]

20. Casbas-Hernandez P, D'Arcy M, Roman-Perez E, Brauer H, McNaughton K, Miller S, Chhetri R, Oldenburg A, Fleming J, Amos K, Makowski L, Troester M. Role of hgf in epithelial-stromal cell interactions during progression from benign breast disease to ductal carcinoma in situ. *Breast Cancer Research*. 15:R82.
21. Sundaram S, Freermerman A, Johnson A, Milner JJ, McNaughton K, Galanko J, Bendt K, Darr D, Perou C, Troester M, Makowski L. Role of hgf in obesity-associated tumorigenesis: C3(1)-tag mice as a model for human basal-like breast cancer. *Breast Cancer Research and Treatment*. 142:489–503. [PubMed: 24218051]
22. Garcia-Closas M, B LA, Lissowska J, Chatterjee N, Peplonska B, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Blair A, Kalaylioglu Z, Rymkiewicz G, Mazepa-Sikora D, Kordek R, Lukaszek S, Sherman ME. Established breast cancer risk factors by clinically important tumour characteristics. *Br J Cancer*. 2006; 95:123–129. [PubMed: 16755295]
23. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61–70. [PubMed: 23000897]
24. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*. 2009; 27:1160–1167. [PubMed: 19204204]
25. Prat A, Parker J, Karginova O, Fan C, Livasy C, Herschkowitz J, He X, Perou C. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research*. 2010; 12:R68. [PubMed: 20813035]
26. Sun X, Gierach GL, Sandhu R, Williams T, Midkiff BR, Lissowska J, Wesolowska E, Boyd NF, Johnson NB, Figueroa JD, Sherman ME, Troester MA. Relationship of mammographic density and gene expression: Analysis of normal breast tissue surrounding breast cancer. *Clinical Cancer Research*.
27. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*. 2001; 98:5116–5121.
28. Elloumi F, Hu Z, Li Y, Parker J, Gulley M, Amos K, Troester M. Systematic bias in genomic classification due to contaminating non-neoplastic tissue in breast tumor samples. *BMC Medical Genomics*. 2011; 4:54. [PubMed: 21718502]
29. Creighton CJ, Casa A, Lazard Z, Huang S, Tsimelzon A, Hilsenbeck SG, Osborne CK, Lee AV. Insulin-like growth factor-i activates gene transcription programs strongly associated with poor breast cancer prognosis. *Journal of Clinical Oncology*. 2008; 26:4078–4085. [PubMed: 18757322]
30. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *Journal of Clinical Oncology*. 2006; 24:1656–1664. [PubMed: 16505416]
31. Troester MA, Hoadley KA, Sørli T, Herbert BS, Børresen-Dale AL, Lønning PE, Shay JW, Kaufmann WK, Perou CM. Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Research*. 2004; 64:4218–4226. [PubMed: 15205334]
32. Proia DA, Kuperwasser C. Reconstruction of human mammary tissues in a mouse model. *Nat Protocols*. 2006; 1:206–214.
33. Buess M, Nuyten D, Hastie T, Nielsen T, Pesich R, Brown P. Characterization of heterotypic interaction effects in vitro to deconvolute global gene expression profiles in cancer. *Genome Biology*. 2007; 8:R191. [PubMed: 17868458]
34. Buess M, R M, Vogel-Durrer BM, Herrmann R, Rochlitz C. Tumor-endothelial interaction links the cd44(+)/cd24(-) phenotype with poor prognosis in early-stage breast cancer. *Neoplasia*. 2009; 10:987–1002. [PubMed: 19794958]
35. Yang X, Figueroa J, Hewitt S, Falk R, Pfeiffer R, Lissowska J, Peplonska B, Brinton L, Garcia-Closas M, Sherman M. Estrogen receptor and progesterone receptor expression in normal terminal duct lobular units surrounding invasive breast cancer. *Breast Cancer Research and Treatment*. 2013:1–11.
36. Meyers M, Klauber-DeMore N, Ollila D, Amos K, Moore D, Drobish A, Burrows E, Dees EC, Carey L. Impact of breast cancer molecular subtypes on locoregional recurrence in patients treated

with neoadjuvant chemotherapy for locally advanced breast cancer. *Annals of Surgical Oncology*. 18:2851–2857. [PubMed: 21442348]

37. Stewart DA, Yang Y, Makowski L, Troester MA. Basal-like breast cancer cells induce phenotypic and genomic changes in macrophages. *Molecular Cancer Research*. 2012; 10:727–738. [PubMed: 22532586]
38. Sansone P, Storci G, Tavoroli S, Guarnieri T, Giovannini C, Taffurelli M, Ceccarelli C, Santini D, Paterini P, Marcu KB, Chieco P, Bonaf xE Massimiliano. Il-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. *The Journal of Clinical Investigation*. 2007; 117:3988–4002. [PubMed: 18060036]
39. Campbell M, Tonlaar N, Garwood E, Huo D, Moore D, Khramtsov A, Au A, Baehner F, Chen Y, Malaka D, Lin A, Adeyanju O, Li S, Gong C, McGrath M, Olopade O, Esserman L. Proliferating macrophages associated with high grade, hormone receptor negative breast cancer and poor clinical outcome. *Breast Cancer Research and Treatment*. 2011; 128:703–711. [PubMed: 20842526]
40. DeNardo DG, Barreto JB, Andreu P, Vasquez L, Tawfik D, Kolhatkar N, Coussens LM. Cd4+ t cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages. *Cancer cell*. 2009; 16:91–102. [PubMed: 19647220]
41. Khan S, Rogers MA, Obando JA, Tamsen A. Estrogen receptor expression of benign epithelium and its association with breast cancer. *Cancer Research*. 1994; 54(4):993–7. [PubMed: 8313390]
42. Khan S, Rogers MA, Khurana KK, Numann PJ. Estrogen receptor expression in benign breast epithelium and breast cancer risk. *J Nat Cancer Inst*. 1998; 90(1):37–42. [PubMed: 9428781]
43. Largent JA, Capanu M, Bernstein L, Langholz B, MellekjÄr L, Malone KE, Begg CB, Haile RW, Lynch CF, Anton-Culver H, Wolitzer A, Bernstein JL. Reproductive history and risk of second primary breast cancer: The wecare study. *Cancer Epidemiology Biomarkers & Prevention*. 2007; 16:906–911.
44. Begg CB, Berwick M. A note on the estimation of relative risks of rare genetic susceptibility markers. *Cancer Epidemiology Biomarkers & Prevention*. 1997; 6:99–103.
45. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi D. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Research*. 2009; 11:R7. [PubMed: 19187537]
46. Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H, Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers WR, Polyak K. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer cell*. 2004; 6:17–32. [PubMed: 15261139]
47. Haslam S, Woodward T. Host microenvironment in breast cancer development: Epithelial-cell-stromal-cell interactions and steroid hormone action in normal and cancerous mammary gland. *Breast Cancer Res*. 2003; 5:208–215. [PubMed: 12817994]
48. Luciani MG, Seok J, Sayeed A, Champion S, Goodson WH, Jeffrey SS, Xiao W, Mindrinos M, Davis RW, Dairkee SH. Distinctive responsiveness to stromal signaling accompanies histologic grade programming of cancer cells. *PLoS ONE*. 2011; 6:e20016. [PubMed: 21625507]
49. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi JT, Rijn Mvd, Botstein D, Brown PO. Gene expression signature of fibroblast serum response predicts human cancer progression: Similarities between tumors and wounds. *PLoS Biol*. 2004; 2:e7. [PubMed: 14737219]

Abbreviations

BC	Breast Cancer
TN	Triple Negative
LumA	Luminal A Breast Cancer
LumB	Luminal B Breast Cancer

HER2	HER2-like Breast Cancer
BLBC	Basal-like Breast Cancer
EReS	Estrogen Response Signature
BCT	Breast Conserving Therapy
ER	Estrogen Receptor
PR	Progesterone Receptor
PWBCS	Polish Women's Breast Cancer Study
NBS	Normal Breast Study
PAM	Predictive Analysis of Microarray
SAM	Significant analysis of Microarray
IPA	Ingenuity Pathway Analysis
RMF	Reduction Mammary Fibroblasts
H&E	Hematoxinilin and Eosin
TDLU	Terminal Ductal Lobular Units
IHC	Immunohistochemistry

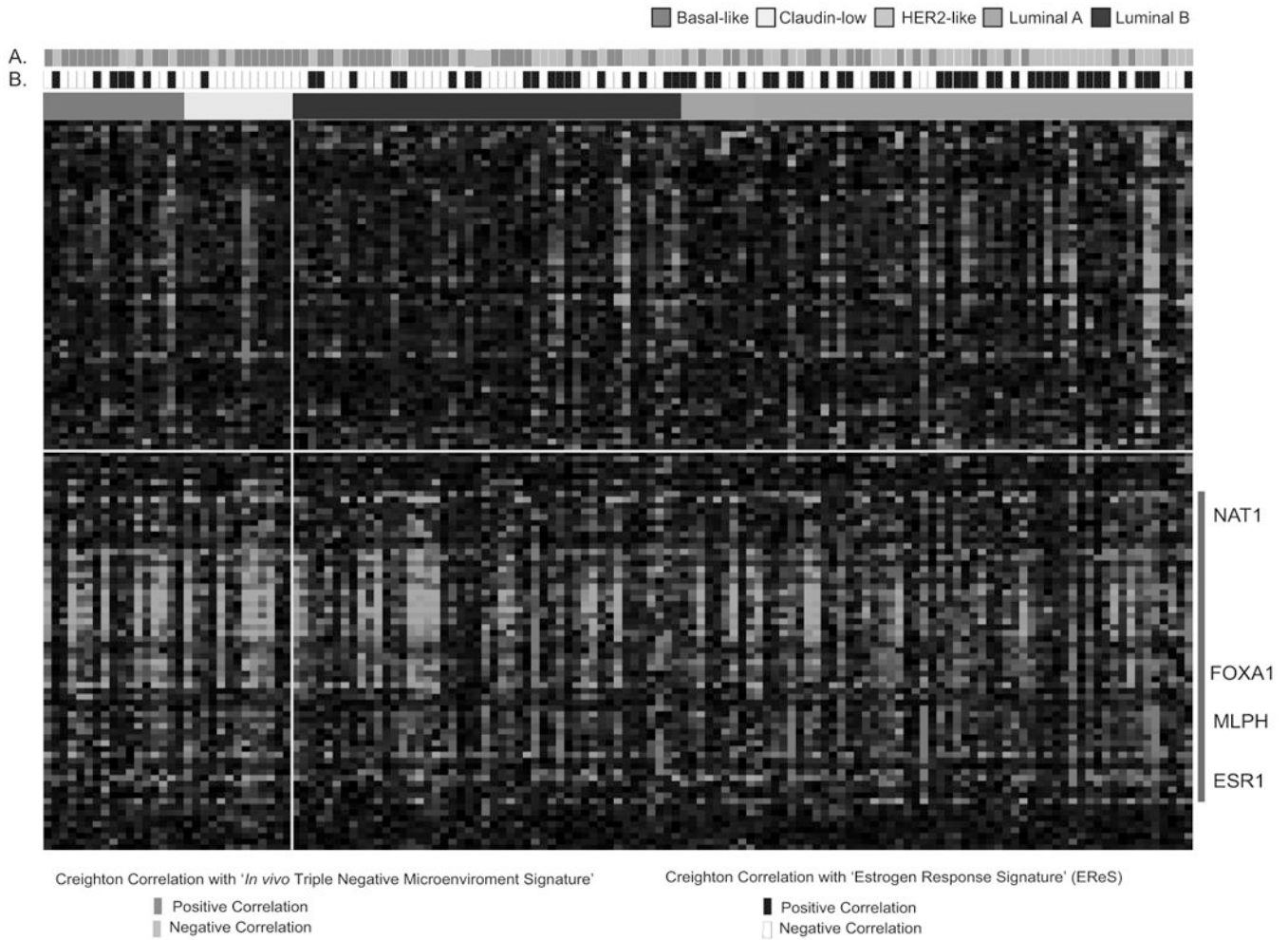


Figure 1. Tumor intrinsic subtype is reflected in cancer-adjacent tissue

Heat map representing 126 differentially expressed genes across the cancer-adjacent tissue of Claudin-low (triple negative), Basal-like (triple negative), Her-2-like, Luminal B and Luminal A in the (PWBCS). Distinct clusters of up and down regulated genes show a trend in the gene expression of these cancer-adjacent tissues. Among those 126 genes, 4 genes that are used in the PAM50 classification system to identify Luminal tumors are also highly expressed in their adjacent tissue. A) Samples were dichotomized as having a positive (black) or negative (white) correlation with the Estrogen response signature (EReS) developed by Oh *et al.* [30] Luminal cancers had the highest proportion of EReS positive cancer-adjacent tissues. B) Samples were dichotomized as having a positive (dark orange) or negative (light orange) correlation with the newly identified 'in vivo triple negative microenvironment signature'. The triple negative cancers had the highest proportion of samples with high expression of this signature.

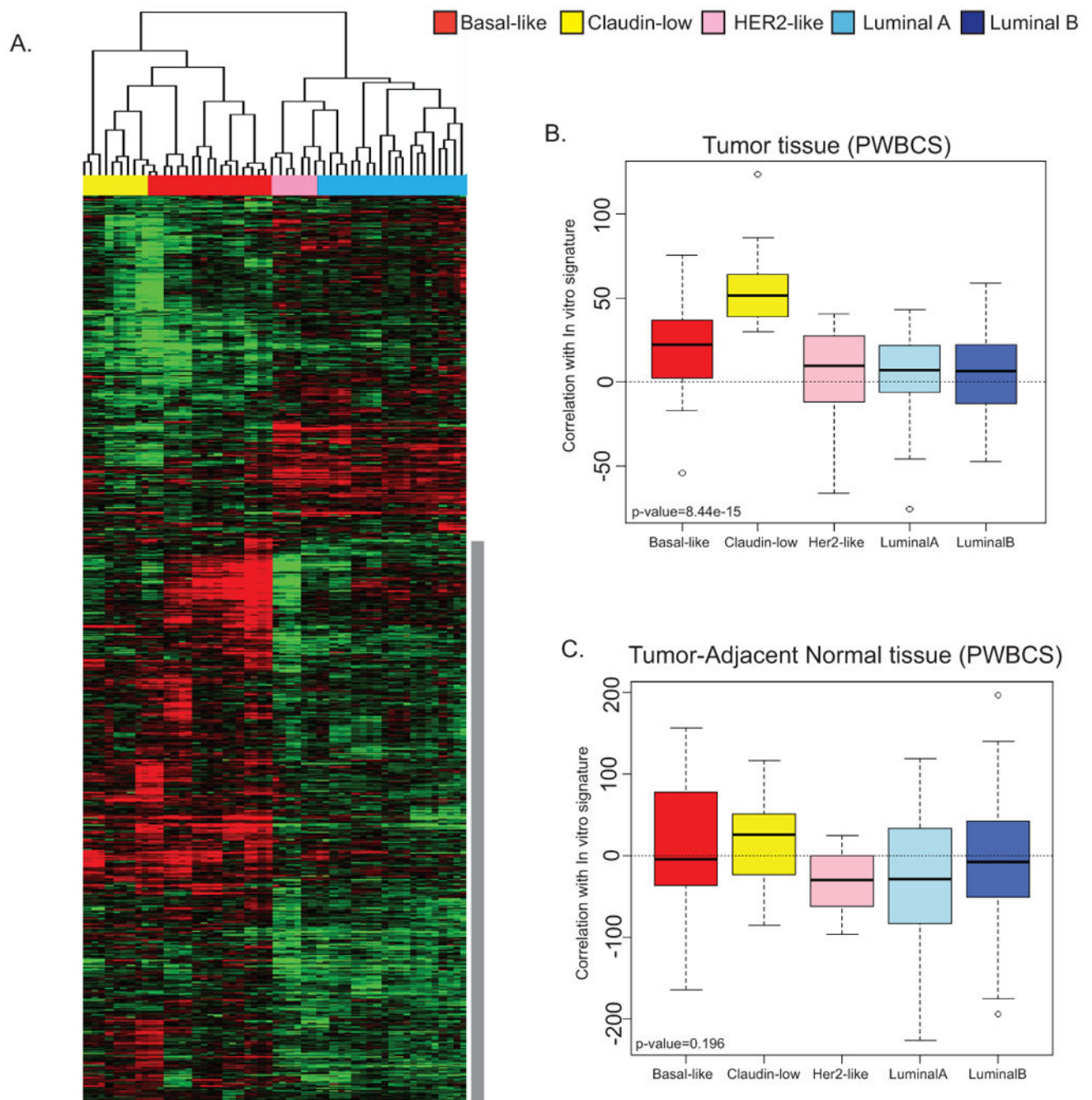


Figure 2. Triple negative microenvironments can be recapitulated in vitro using coculture systems

(A) Direct *in vitro* cocultures were performed by culturing fibroblasts with cell lines models of different subtypes of breast cancer. Heat map represents interaction values (I) from the Buess et al. [33] method of identifying genes differentially expressed in triple negatives versus all the other cocultures. The grey bar represents genes upregulated in triple negative direct cocultures ("In vitro triple negative signature"). This *in vitro* generated list of genes ("In vitro triple negative signature", grey bar in A), is highly expressed both in the

intratumoral microenvironment (**B** – Tumor tissue) and in the cancer-adjacent tissue (**C** – Tumor adjacent normal tissue).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Expression of the estrogen response gene signature (EReS), and the 'In vivo triple negative microenvironment signature' according to tumor subtype in PWBCS and TCGA

	PWBCS N (%)						TCGA N (%)					
	Basal-like	Claudin-Low	LumB	LumA	Fishers Exact p-value	Basal-like	Claudin-Low	LumB	LumA	Fishers Exact p-value		
EReS (+)	7 (41.2)	1 (7.7)	19 (40.4)	33 (62.3)	0.004	5 (55.6)	-	9 (75.0)	17 (68.0)	0.001		
EReS (-)	10 (58.3)	12 (92.3)	28 (59.6)	20 (37.7)		4 (44.4)	10 (100.0)	3 (25.0)	8 (32.0)			
In vivo TN sig (+)	12 (70.6)	11 (84.6)	27 (57.4)	18 (34.0)	0.003	3 (33.3)	10 (100.0)	6 (50.0)	7 (28.0)	0.001		
In vivo TN sig (-)	5 (29.4)	2 (15.4)	20 (42.6)	35 (66.0)		6 (66.7)	-	6 (50.0)	18 (72.0)			

Table 2

Cancer-adjacent tissue composition according to tumor subtype and estrogen response gene signature (EReS) in PWBCS and TCGA.

	Basal-like n=17	Claudin-low n=11	LumB n=47	LumA n=52	Fisher's Exact p-value	EReS (+) n=59	EReS (+) n=68	χ^2 square p-value
PWBCS N (%)								
% Epithelium^a								
<10%	13 (76.5)	4 (36.3)	29 (61.7)	33 (63.5)	0.213	37 (62.7)	42 (61.8)	1.000
>=10%	4 (23.5)	7 (63.6)	18 (38.8)	19 (36.5)		22 (37.3)	26 (38.2)	
% Stroma^b								
<20%	9 (52.9)	8 (72.7)	24 (51.7)	23 (44.2)	0.235	12 (20.3)	52 (76.5)	<0.001
>=20%	8 (47.1)	3 (27.3)	23 (48.9)	29 (55.7)		47 (79.7)	16 (23.5)	
TCGAN N (%)								
% Epithelium^a								
<10%	6 (75.0)	8 (88.9)	5 (50.0)	10 (50.0)	0.177	11 (47.8)	18 (75.0)	0.075
>=10%	2 (25.0)	1 (11.1)	5 (50.0)	10 (50.0)		12 (52.2)	6 (25.0)	
% Stroma^b								
<20%	7 (87.5)	9 (100.0)	7 (90.0)	12 (60.0)	0.104	14 (60.9)	21 (87.5)	0.049
>=20%	1 (12.5)	-	3 (30.0)	8 (40.0)		9 (39.1)	3 (12.5)	

^a average % epithelium=9.8%^b average % non-fatty stroma=26.8%.