Benign Breast Tissue Composition in Breast Cancer Patients: Association with Risk Factors, Clinical Variables, and Gene Expression

Xuezheng Sun1, Rupninder Sandhu2, Jonine D. Figueroa3, Gretchen L. Gierach3, Mark E. Sherman3,4, and Melissa A. Troester1,2

Abstract

Background: Breast tissue composition (epithelium, non-fatty stroma, and adipose) changes qualitatively and quantitatively throughout the lifespan, and may mediate relationships between risk factors and breast cancer initiation. We sought to identify relationships between tissue composition, risk factors, tumor characteristics, and gene expression.

Methods: Participants were 146 patients from the Polish Breast Cancer Study, with data on risk factor and clinicopathological characteristics. Benign breast tissue composition was evaluated using digital image analysis of histologic sections. Whole-genome microarrays were performed on the same tissue blocks.

Results: Mean epithelial, non-fatty stromal, and adipose proportions were 8.4% (SD = 4.9%), 27.7% (SD = 24.0%), and 64.0% (SD = 24.0%), respectively. Among women <50 years old, stroma proportion decreased and adipose proportion increased with age, with approximately 2% difference per year (P < 0.01). The variation in epithelial proportion with age was modest (0.1% per year). Higher epithelial proportion was associated with obesity (7.6% in nonobese vs. 10.1% in obese; P = 0.02) and with poorly differentiated tumors (7.8% in well/moderate vs. 9.9% in poor; P = 0.05). Gene expression signatures associated with epithelial and stromal proportion were identified and validated. Stroma-associated genes were in metabolism and stem cell maintenance pathways, whereas epithelial genes were enriched for cytokine and immune response pathways.

Conclusions: Breast tissue composition was associated with age, body mass index, and tumor grade, with consequences for breast gene expression.

Impact: Breast tissue morphologic factors may influence breast cancer etiology. Composition and gene expression may act as biomarkers of breast cancer risk and progression.

Introduction

Benign breast tissue is composed of epithelium, stroma, and adipose tissue, each of which is heterogeneous and complex in composition. Previous papers have suggested that during critical developmental windows, such as embryogenesis, pregnancy, and postmenopausal involution, important shifts occur in normal breast tissue composition (1, 2). Epithelial proportions increase with hormonal exposure, and stromal composition may also increase (3-5). As malignant transformation evolves, composition of benign breast tissues may change in parallel (6-9), exerting selective forces that influence tumor biology and behavior (10, 11). It is important to understand how the balance of epithelium and stroma is modulated by normal events throughout the lifespan, and how this balance contributes to carcinogenesis. The importance of understanding factors that influence benign breast tissue composition is underscored by the strong direct association between mammographic density (a radiological measure of fibroglandular breast tissue content) and breast cancer risk and the growing experimental evidence that the tumor microenvironment may influence breast cancer biology, prognosis, and treatment response (12-15).
Breast Tissue Composition

The epidemiologic and clinical factors that affect benign breast tissue can be partly understood by examining mammographic density. Specifically, increasing age, menopause, elevated body mass index (BMI), and having a live birth are associated with lower mammographic density, suggesting less fibroglandular tissue relative to adipose tissue. Family history of breast cancer and use of menopausal hormones are related to higher mammographic density, and therefore greater fibroglandular content. In benign breast tissue of younger women, stroma accounts for the majority of mammographically dense tissue, but epithelial content may be an important driver of mammographic density-risk associations. Few studies have directly parsed density into stroma versus epithelium, and even fewer have evaluated all three components (epithelium, fat, and stroma) in association with risk factors (16–18). Such studies could lead to improved understanding of mammographic density as a risk factor and may identify novel risk biomarkers.

To determine factors related to breast histology and to identify plausible biologic pathways affected by changing breast tissue composition, we used breast tissues and gene expression data from the Polish Breast Cancer Study. Our aims were to determine the relationships of breast tissue composition to breast cancer risk factors and tumor characteristics, and to relate different tissue components (epithelial, non-fatty stroma, and adipose) to gene expression profiles.

Materials and Methods

Patient samples
The study population is a subset of patients from the Polish Breast Cancer Study (PBCS), a population-based case–control study conducted in two large cities in Poland (Warsaw and Łódź) during 2000–2003 (19). PBCS cases were women ages 20 to 74 years with an incident, pathologically confirmed in situ or invasive breast carcinoma identified through a rapid identification system organized at five participating hospitals and via cancer registries. Fresh tissues from invasive tumors, non-neoplastic adjacent breast tissue, and mammary fat tissue were collected at the time of breast surgery and snap frozen in liquid nitrogen. On the basis of in vitro evidence that basal-like and luminal tumors are associated with distinctive micro-environments (20), we oversampled these subtypes. All of the participants provided written informed consent under a protocol approved by the U.S. National Cancer Institute (Bethesda, MD) and local (Polish) institutional review boards.

We limited our analysis to 146 invasive breast cancer cases in Warsaw with available data on breast composition and gene expression of extratumoral breast tissues. Our interest was to evaluate how normal tissue without evidence of pathologic change may vary in association with risk factors. Considering that epithelial tissue is a minority component (typically <10% by area) in normal breast (16), 11 women with outlying values suggestive of possible hyperplasia (epithelial proportion more than 20%) were excluded as potentially non-normal. We further excluded women with incomplete data on breast risk factors and clinicopathological factors, which left 96 women in our main analysis. The 50 women with incomplete risk factor data were included as a validation dataset.

Data collection
Information on demographic and anthropometric factors was collected from medical records and in-person interviews (19). Mammographic density of pretreatment mammograms of the unaffected breast was assessed using a quantitative, reliable computer-assisted thresholding method (21). One expert reader measured absolute dense area (cm²) and total breast area (cm²); percentage mammographic density was calculated by dividing the dense breast area by the total breast area and multiplying by 100. Frozen non-neoplastic breast specimens of approximately 100 mg were cut over dry ice and then used to prepare 20 µm frozen sections from each end. The central portion was used for RNA extraction. Microarrays on the central section were performed using two-color 4 x 44K Agilent whole-genome arrays. More information on RNA isolation, microarray, and data preprocessing has also been described previously (18). Microarray data used in this analysis are publicly available through the Gene Expression Omnibus (GSE49175).

After hematoxylin and eosin (H&E) staining, the sections were scanned into high-resolution (20×) digital images using the Aperio Scan-Scancope XT Slide Scanner (Aperio Technologies) in the Translational Pathology Laboratory of University of North Carolina at Chapel Hill. The details of composition measurement and measurement validation have been described previously (18). Briefly, 15 representative digital slides were first manually annotated for epithelial area, non-fatty stromal area, and total area (mm²) using Aperio ImageScope software. These annotations were used to train the composition estimator in Aperio’s Genie Classifier. The trained classifier was positively and strongly correlated with manual annotation, and showed better performance in quantifying small percentages over visual assessment (by human eye on regular H&E slides). After training and validation, Aperio’s Genie Classifier was applied to the remaining slides partitioning epithelium, adipose tissue, non-fatty stroma, and glass into percentages.

Statistical analysis
The distribution of adipose, non-fatty stroma, and epithelial composition and the Pearson correlation between components were evaluated using R (version 3.0.0). To assess the influence of age on breast composition, we evaluated the distribution of composition by five age groups at diagnosis (<40 years, 40–49 years, 50–59 years,
60–69 years, and ≥70 years) and used a cubic spline interpolation algorithm (knots = 3) to model and visualize these associations (22). As stromal and adipose proportions were strongly inversely correlated (γ = −0.98, P < 0.001), we only studied epithelial and stromal compositions in subsequent analyses. To study how composition changed in one compartment (e.g., epithelium) relative to another (e.g., stroma), the ratios of epithelial-to-stromal proportions and stromal-to-adipose proportions were also evaluated. Log-transformed ratios were used in the analysis to account for substantial skew.

Besides age, other breast cancer risk factors included percentage density, dense area, BMI, use of oral menopausal hormone therapy, menopausal status, family history of breast cancer in a first-degree relative, previous benign breast disease history (defined as ever having any biopsy or partial removal procedure, with no cancer detected), age at menarche, and parity. Age at first birth and duration of lactation were assessed among parous women. Time since last birth could not be evaluated because approximately 95% patients had their most recent birth more than 15 years before breast cancer diagnosis. Tumor clinicopathological factors evaluated included tumor size, histological type, differentiation stage, number of positive axillary lymph nodes, and ER status. The age-adjusted differences in epithelial and stromal compositions according to breast cancer risk factors and tumor characteristics were compared using Wald test.

Linear regression with Linear Models for Microarray Data (23) was used to identify genes associated with stromal and epithelial composition. Both epithelial and stromal proportions were included in the model as independent variables. The false discovery rate (FDR) approach was used to control for multiple comparisons, with 0.05 as the cutoff for statistical significance. Previous analyses showed distinct microenvironments associated with breast tumor subtypes (20, 24), so ER-stratified analyses showed distinct microenvironments associated with breast tumor subtypes. The FDR approach was used for multiple comparisons, with 0.05 as the cutoff for statistical significance.

To validate the signatures identified, a test dataset (n = 50, described above) was used. Principle component analysis was used to assess global gene expression in the test and training datasets, and showed considerable overlap (Supplementary Fig. S1). To evaluate relative expression of a signature in a test set sample, the log2(R/G) expression for each gene was multiplied by corresponding direction index (1 for upregulated genes, and -1 for downregulated gene) and results were summed across all genes. This score was also evaluated for association with cellular composition of test samples. Although training and test datasets showed considerable similarity in demographic, anthropometric factors, and tumor characteristics (except that mammographic density of women in test dataset was slightly higher compared with women in training dataset), we performed a sensitivity analysis including the 50 subjects, to evaluate the impact of exclusion from the training dataset due to missing values.

The magnitude of associations observed in the training dataset was not meaningfully changed by including these patients.

Results

Associations between histology and risk factors/clinicopathological characteristics

The analysis used histology data from 96 patients with complete data to assess risk factor associations. In this sample, mean epithelial proportion was 8.4% (SD = 4.9%; range = 0.5%–19.4%), mean stromal proportion was 27.7% (SD = 24.0%; range = 0.02%–85.0%), and mean adipose proportion was 64.0% (SD = 24.0%; range = 12.5%–98.5%). Although epithelial proportion was independent of the other two components (non-fatty stroma: γ = −0.09, P = 0.40; adiposity: γ = −0.12, P = 0.26), stromal and adipose proportions were strongly inversely correlated (γ = −0.98, P < 0.001).

Among the risk factors examined, age showed the strongest association with breast composition. Interestingly, the association was not strictly monotonic linear, but had a turning point around the age of 50 years (Fig. 1). Adipose proportion significantly increased with age until the age of 50 years, with a 2.0% increment per year (P = 0.004). After the age of 50 years, the rate of change in adipose proportion decreased to 0.1% per year (P = 0.800). Conversely, stromal proportion decreased with age at a rate of −1.9% per year (P = 0.006), with an inflection after 50 years of age, and a rate of −0.06% per year (P = 0.864) thereafter. The variation in epithelial composition with age was relatively small, both before and after the age of 50 years, but a slight U-shape curve was observed, with declines in epithelial proportion before the age of 50 years, followed by a modest increase in proportion thereafter (< 50 years: 0.11% decrement per year, P = 0.426; ≥50 years: 0.14% increment per year, P = 0.042). The ratios of epithelium to non-fatty stroma and non-fatty stroma to adipose were also assessed (Supplementary Fig. S2). For the ratio of epithelium to non-fatty stroma, we observed a slight but insignificant increase with age. The ratio of stromal-to-adipose proportion had an association with age that was very similar to that of stromal proportion, decreasing till the age of 50 years and then leveling off.

The high inverse correlation between stromal and adipose proportions, subsequent analyses focused on epithelium and non-fatty stroma, and did not separately assess associations with adipose proportion. However, the adipose associations are presented in Supplementary Table S1. Factors associated with breast composition after adjustment for age, included BMI and mammographic density measures (Table 1). We evaluated potential confounding of risk factor–composition relationships by BMI, and found no substantial confounding except for in models with mammographic density, and therefore reported the parsimonious model without BMI for all other associations. Women with higher stromal proportion were more likely to have higher percentage density and dense...
area, but were less likely to be obese. Interestingly, epithelial composition showed a positive association with BMI, but was suggested to be inversely associated with breast density. Although not statistically significant, stromal but not epithelial content, was lower in parous compared with nulliparous women ($P = 0.06$). We did not detect any significant associations of breast composition with tumor characteristics, except that higher epithelial proportion was associated with poorly differentiated tumors (Table 2).

Association between histology and gene expression

To identify gene expression that varied with breast composition on 96 women in the training sample with complete risk factor data, supervised analyses were performed. We detected an epithelial-content–associated signature ($n = 716$ genes, FDR $< 0.05$ and $\beta > 0.02$), with approximately half of the genes upregulated ($n = 332$; Supplementary Table S2). Pathway analysis showed that upregulated genes included cytokine/chemokine activity, EIF2 signaling, and mTOR signaling, and suggested biologic functions related to epithelial regulation and immune-related processes (Supplementary Table S3). No significant pathways (defined as Benjamini–Hochberg adjusted $P < 0.05$) were detected among the downregulated genes.

Stromal composition was associated with a broad biologic profile, with the corresponding signature including 1,065 genes (484 downregulated and 581 upregulated genes) defined by FDR $< 0.05$ and $\beta > 0.02$ (Supplementary Table S2). The upregulated pathways were involved in regulating vascular permeability, epithelial-to-mesenchymal transition, and maintenance of breast cancer stem cells in self-renewal state (Supplementary Table S3). The downregulated pathways included metabolic and immune-related signaling pathways. The stromal and epithelial signatures were relatively independent, with less than one third of genes being identified in both signatures and only cell-sertoli cell junction signaling pathways being identified in both gene sets.

Given previous evidence of expression differences (20, 24–26) and our own findings suggesting epithelial content differences in cancer-adjacent normal by tumor ER status (27), we also performed analysis of gene expression stratified by ER status. In ER-positive patients, the stromal and epithelial signatures were similar to the signatures detected in unstratified analysis, with considerable overlap (stromal signature, 97%; epithelial signature, 65%). In ER-negative patients, no epithelial signature was detected. However, ER-negative stroma had a unique signature, with 425 genes identified (up $= 190$; down $= 235$; Supplementary Table S2) that were not captured in unstratified/overall analysis. Pathway analysis showed that these genes are involved in carcinogenesis and tumorigenesis, for example, actin cytoskeletal signaling, Wnt/β-catenin signaling, and leukocyte extravasation signaling (Supplementary Table S3).

We validated the composition-associated signatures in an independent subsample data ($n = 50$) from the PBCS. As shown in Fig. 2A, the average gene expression level of

Figure 1. Boxplots and spline curves for breast tissue composition by age. Boxplots show the distribution of composition by five age groups at diagnosis (<40 years, 40–49 years, 50–59 years, 60–69 years, and >70 years). A cubic spline interpolation algorithm (knots = 3) was used to smooth the curves and visualize these associations.
our stromal content signature increased with actual stromal proportion ($P_{\text{trend}} < 0.001$, age-adjusted). The high inverse correlation between non-fatty stroma and adipose compositions was also reflected at the gene expression level, where the expression of the stromal signature decreased with actual adipose proportion (Fig. 2B, $P_{\text{trend}} < 0.001$, age-adjusted). Similarly, expression level of epithelial signature showed positive associations with epithelial proportion (Fig. 2C, $P_{\text{trend}} < 0.001$, age-adjusted).

Thus, the signatures identified in the training dataset showed strong relationships with composition in the independent validation data.

### Discussion

Benign breast tissue composition is responsive to exogenous exposures, endogenous physiologic effects, and pathologic events. Understanding the factors that influence composition is important for understanding breast cancer etiology, risk, and progression. Our results showed that epithelial and stromal proportions were associated with well-recognized breast cancer risk factors (such as age, BMI, and mammographic density) and with tumor differentiation. The strongest associations were between age and breast composition. Although the effect of age on breast epithelial content is well recognized (16, 17, 28, 29),

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Epithelial percentage</th>
<th>Stromal percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>Mean</td>
</tr>
<tr>
<td>Percentage density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25%</td>
<td>50 (52)</td>
<td>8.69</td>
</tr>
<tr>
<td>≥25%</td>
<td>46 (48)</td>
<td>7.99</td>
</tr>
<tr>
<td>Dense area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;Median&lt;sup&gt;b&lt;/sup&gt;</td>
<td>52 (54)</td>
<td>8.25</td>
</tr>
<tr>
<td>≥Median</td>
<td>44 (46)</td>
<td>8.48</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30 kg/m&lt;sup&gt;2&lt;/sup&gt;</td>
<td>67 (70)</td>
<td>7.59</td>
</tr>
<tr>
<td>≥30 kg/m&lt;sup&gt;2&lt;/sup&gt;</td>
<td>29 (30)</td>
<td>10.12</td>
</tr>
<tr>
<td>Menopausal status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>23 (24)</td>
<td>9.22</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>73 (76)</td>
<td>8.08</td>
</tr>
<tr>
<td>Age at menarche</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤12 y</td>
<td>38 (40)</td>
<td>8.36</td>
</tr>
<tr>
<td>&gt;12 y</td>
<td>58 (60)</td>
<td>8.38</td>
</tr>
<tr>
<td>Parity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>20 (21)</td>
<td>7.62</td>
</tr>
<tr>
<td>≥1</td>
<td>76 (79)</td>
<td>8.55</td>
</tr>
<tr>
<td>Age at first birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤25 y</td>
<td>45 (59)</td>
<td>9.15</td>
</tr>
<tr>
<td>&gt;25 y</td>
<td>31 (41)</td>
<td>7.78</td>
</tr>
<tr>
<td>Duration of lactation among parous women</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6 m</td>
<td>32 (53)</td>
<td>9.43</td>
</tr>
<tr>
<td>&gt;6 m</td>
<td>28 (47)</td>
<td>8.67</td>
</tr>
<tr>
<td>MHT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>64 (69)</td>
<td>7.90</td>
</tr>
<tr>
<td>Current/ever</td>
<td>29 (31)</td>
<td>9.17</td>
</tr>
<tr>
<td>Previous benign breast disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>73 (78)</td>
<td>8.07</td>
</tr>
<tr>
<td>Yes</td>
<td>20 (22)</td>
<td>9.38</td>
</tr>
<tr>
<td>Family history&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>77 (81)</td>
<td>8.12</td>
</tr>
<tr>
<td>Yes</td>
<td>18 (19)</td>
<td>9.60</td>
</tr>
</tbody>
</table>

Abbreviation: MHT, menopausal hormone therapy; SE, standard error.

<sup>a</sup>Age was adjusted as linear spline with the knot of 50 years. For percentage density and dense area, BMI was additionally adjusted.
<sup>b</sup>P values were calculated using Wald test.
<sup>c</sup>Median = 36.52 cm<sup>2</sup>.
<sup>d</sup>Family history of breast cancer in first-degree relative.
we also observed significant differences in stromal content. The association between age and tissue composition mirrors associations with percentage density (30), with neither appearing strictly monotonic, but both suggesting a turning point at the age of 50 years.

The observation that stroma is slowly displaced by fat even before menopause has potential implications for breast tumorigenesis broadly and for epidemiologic studies of breast cancer risk. It is increasingly recognized that stroma plays an important role in breast cancer progression (31–33). It has also been shown that expression of adipokines may be altered with obesity and in the peritumoral tissue of advanced breast tumors, suggesting a role for breast fat and adipokines in tumorigenesis (34).

However, the impact of changing premenopausal breast tissue composition on risk is largely unexplored. Given the degree of changes in the breast microenvironment that occur before menopause, future studies should consider stromal composition as a possible modifier of risk among premenopausal women. Measures of microenvironment composition are not widely available; however, mammographic density is increasingly available in epidemiologic studies and may be not only a surrogate marker (35), but also a risk modifier for some carcinogenic exposures.

It is interesting that the premenopausal stromal changes were more pronounced in our study than the age-related changes in epithelium. In contrast with previous findings (16), we did not observe statistically significant age-associated differences in epithelial proportion among younger (age < 50 years) women. Differences in the age-epithelial content association between our study and Nurse’s Health Study’s (NHS) benign breast disease cohort (16) may be due to the older age of women in our dataset (percentage of population < 50 years: 22% vs. 68% in NHS). That is, given that our study included fewer younger women (where the effects of age are most pronounced), we were underpowered to detect age-related declines in epithelial content. However, our data including more older women yielded interesting novel insights, including suggestions of a modest increase in epithelial proportion after the age of 50 years. As women age, terminal duct lobular units (TDLU) are well known to regress, likely due to hormonal changes; however, large ducts are believed to change less under the influence of age (27, 36). This could account for nonlinearity in the age-epithelial content association. These associations should be further investigated in other cohorts, particularly if epithelial content serves as a marker of risk as suggested previously (28).

Composition changes with age may also reflect changes in breast volume. As non-fatty stroma declines, overall breast tissue volume declines and there is a concomitant increase in fat proportion. This is reflected in the ratio of non-fatty stroma to adipose: the ratio increased slightly in the age group of 60 to 69 years, decreasing after the age of 70 years. Thus, this represents an alternative explanation for the increase in epithelial content in older women; the increasing proportion of epithelium may reflect a smaller breast rather than an increase in epithelium. However, even after normalizing epithelial proportion to fat percentage, the U-shaped curve morphology persisted (data not shown). Strong dichotomies in breast composition between young women (<50) and older women (≥50) are important to consider in light of observations that tumor biology in young women differs from that of old women (29).

BMI was strongly and inversely associated with non-fatty stroma content in PBCS. This result was consistent with previous studies (16, 17), perhaps reflecting the

| Table 2. Age-adjusted means of epithelial and stromal proportion by tumor characteristics |
|---------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Characteristics                            | N (%) | Mean (%) | SE   | P    | Mean (%) | SE   | P    |
| Tumor size                                 |       |          |      |      |          |      |      |
| < 2 cm                                     | 51 (53)| 8.11     | 0.67 | 0.59 | 30.38    | 3.17 | 0.22 |
| ≥ 2 cm                                     | 45 (47)| 8.64     | 0.71 |      | 24.66    | 3.38 |      |
| Histologic type                            |       |          |      |      |          |      |      |
| Ductal                                     | 51 (53)| 8.05     | 0.67 | 0.51 | 27.81    | 3.19 | 0.96 |
| Other                                      | 45 (47)| 8.70     | 0.71 |      | 27.58    | 3.40 |      |
| Affected lymph node                        |       |          |      |      |          |      |      |
| 0                                          | 70 (73)| 8.33     | 0.66 | 0.96 | 28.62    | 3.17 | 0.67 |
| ≥ 1                                        | 26 (27)| 8.38     | 0.72 |      | 26.61    | 3.44 |      |
| Differentiation                            |       |          |      |      |          |      |      |
| Well/moderate                              | 52 (54)| 7.78     | 0.56 | 0.05 | 29.58    | 2.70 | 0.18 |
| Poor                                       | 44 (46)| 9.91     | 0.92 |      | 22.64    | 4.44 |      |
| ER status                                  |       |          |      |      |          |      |      |
| Negative                                   | 24 (25)| 9.76     | 1.00 | 0.11 | 22.25    | 4.78 | 0.19 |
| Positive                                   | 72 (75)| 7.89     | 0.56 |      | 29.52    | 2.69 |      |

*aAge was adjusted as linear spline with the knot of 50 years. P values were calculated using Wald test.*
coordination of whole body composition and local breast composition. However, obese women also had a higher proportion of epithelium. This association has not been reported previously, but is line with the positive relationship of obesity and breast cancer risk observed in postmenopausal women. Adipose cells are believed to express high levels of aromatase, and therefore local effects on epithelial cell histology are biologically plausible (37).

Because 76% of our study population was postmenopausal, our data were well powered to capture the association of epithelial variation with obesity.

We did not observe strong associations between epithelial content and reproductive factors, in contrast with previous findings in the NHS (16). The vast majority of our patients had experienced birth more than 15 years before sample collection, whereas the younger NHS study participants (mean age 45 years compared with 56 years in PBCS) were more recently pregnant. The differences in findings and study population suggest that effects of reproductive exposures, such as parity, on breast tissue composition may not be persistent. In fact, the prevailing hypothesis for long-term parity-associated protection emphasizes change in differentiation state of epithelium rather than change in cellular content (38). Other morphometric measures of epithelial size, shape, and or involution status may be more closely linked with reproductive factors.

Because patients in this study had cancer, it was also possible to evaluate breast composition in association with tumor characteristics. We found that greater epithelial composition was associated with poorer tumor differentiation and weakly with ER-negative status. Although the effects suggesting relationships between cancer-adjacent composition and tumor characteristics are weak, future studies should consider other measures of benign morphometry. A previous study in PBCS (27) has suggested that TDLU involution may be inversely correlated with basal-like breast cancer and more careful measures of epithelial morphometry may be more sensitive than "epithelial area" used in this study. Future studies should also evaluate inflammatory infiltrates as these cells may play an important role in gene expression.

Figure 2. A–C, the average gene expression level of epithelial and stromal composition signatures across breast tissue composition groups in a test dataset (n = 50). The association of relative expression of epithelial and stromal composition signatures with breast tissue composition was validated in a test set sample (n = 50). The log2(R/G) expression for each gene of each composition signature was multiplied by the corresponding direction index (1 for upregulated genes and -1 for downregulated genes), and results were summed across all genes as the relative gene expression level. This relative gene expression level was evaluated for association with tissue composition by tertiles (cutoff points: epithelial proportion, 7.20% and 16.05%; stromal proportion, 10.78% and 34.31%; adipose proportion, 47.11% and 79.62%). A and C, the average gene expression level of the stromal and epithelial composition signatures increased with actual stromal and epithelial proportion (age-adjusted P_cortrend < 0.001). B, the expression level of the stromal signature decreased with actual adipose proportion (age-adjusted P_cortrend < 0.001).
patterns of ER-positive versus negative tumors. However, an interesting observation of our study is that the associations of epithelium and non-fatty stroma with tumor characteristics were always in opposite directions. For example, large tumor size was associated with decreased stromal content and increased epithelial content. This could reflect a repressive stroma and/or a common factor that increases epithelial cell growth for both benign and tumor tissue. Whatever the causes, these disparate findings for stroma and epithelium have important implications. Because epithelial and stromal contents together are associated with increasing mammographic density, relationships between cellular content and tumor characteristics may be obscured (i.e., changes in epithelium and stroma may effectively cancel each other out) in studies that rely on mammographic density as a surrogate measure of breast tissue composition. Nonetheless, although associations between mammographic density and intrinsic subtype may be largely null (39, 40), associations between tumor characteristics and particular cellular components may be important for progression.

By linking our risk factor exposures with composition, and composition with gene expression, our findings suggest biologic pathways that are altered by risk factor exposure. For example, risk factors that alter stromal content may have effects on vascular permeability, tight junction signaling, metabolic signaling, loss of contact inhibition, and cell polarity. Indeed, our previous experimental work has shown that stromal–epithelial interactions altered metabolism and glucose uptake of breast cancers (41). Thus, these gene expression changes may lead to novel hypotheses about breast cancer microenvironments. Furthermore, these independently validated composition signatures may be used in future studies where histologic data are unavailable, to estimate composition.

Our study should be considered in light of limitations and strengths. The non-neoplastic breast tissues used in the present study were taken from patients with breast cancer. The influence of tumor on extratumoral microenvironment has been reported previously (24, 42), and therefore associations should be interpreted with caution of these possible differences. Future histologic research on normal breast tissues from healthy women is needed. In particular, studies like the Komen Tissue Bank (43), in which needle biopsy sampling is essentially "blind" to tissue structure, are important. In studies using whole tissue collected during pathologic evaluation, a bias toward sampling tissue that seems to contain parenchymal cells is well known. This bias was also present in the PBCS samples. Thus, our findings on composition are generalizable only as relative measures of composition and not as absolute measures. It is likely that our study, and moreover, most studies using surgical tissues, overestimate stromal and epithelial percentages. Despite these limitations, these data show compelling evidence that breast tissue composition varies according to epidemiologic risk factors, with important consequences for gene expression.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: X. Sun, R. Sandhu, J.D. Figueroa, M.E. Sherman, M.A. Troester
Development of methodology: X. Sun, R. Sandhu, M.E. Sherman, M.A. Troester
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Sandhu, J.D. Figueroa, M.E. Sherman, M.A. Troester
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X. Sun, R. Sandhu, J.D. Figueroa, G.L. Gierach, M.E. Sherman, M.A. Troester
Writing, review, and/or revision of the manuscript: X. Sun, R. Sandhu, J.D. Figueroa, G.L. Gierach, M.E. Sherman, M.A. Troester
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X. Sun, R. Sandhu, M.A. Troester
Study supervision: M.A. Troester

Acknowledgments
The authors thank Norman F. Boyd for contributing density measurements used in this analysis and for thoughtful comments on the article and Drs. Montserrat Garcia-Closas, Jolanta Lisowski, and Louise Brinton for their contributions in data and biospecimen collection.

Grant Support
This work was supported by grants from the National Cancer Institute (U01-ES019472 and R01-CA138255 to M.A. Troester), a Breast SPORE Project (grant P50CA050223, H.S. Earp PI), and a grant from the Avon Foundation (to M.A. Troester). This work was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 7, 2014; revised August 22, 2014; accepted August 29, 2014; published OnlineFirst September 23, 2014.

References
8. Kass L, Erief JT, Dembo M, Weaver VM. Mammary epithelial cell: Influence of extracelular matrix composition and organization during...
Benign Breast Tissue Composition in Breast Cancer Patients: Association with Risk Factors, Clinical Variables, and Gene Expression

Xuezheng Sun, Rupinder Sandhu, Jonine D. Figueroa, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-14-0507

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2014/09/24/1055-9965.EPI-14-0507.DC1

Cited articles
This article cites 42 articles, 16 of which you can access for free at:
http://cebp.aacrjournals.org/content/23/12/2810.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/23/12/2810.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.