

RESEARCH ARTICLE

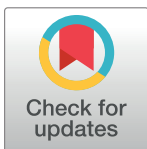
Correlated metabolomic, genomic, and histologic phenotypes in histologically normal breast tissue

Xuezheng Sun¹✉, Delisha A. Stewart²✉, Rupninder Sandhu³, Erin L. Kirk¹, Wimal W. Pathmasiri², Susan L. McRitchie², Robert F. Clark⁴, Melissa A. Troester^{1,3,5}, Susan J. Sumner^{2*}

1 Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **2** National Institutes of Health Eastern Regional Comprehensive Metabolomics Resource Core, Department of Nutrition, University of North Carolina at Chapel Hill, Nutrition Research Institute, Kannapolis, North Carolina, United States of America, **3** Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** Analytical Chemistry and Pharmaceuticals, RTI International, Durham, North Carolina, United States of America, **5** Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

✉ These authors contributed equally to this work.

* susan_sumner@unc.edu



OPEN ACCESS

Citation: Sun X, Stewart DA, Sandhu R, Kirk EL, Pathmasiri WW, McRitchie SL, et al. (2018) Correlated metabolomic, genomic, and histologic phenotypes in histologically normal breast tissue. *PLoS ONE* 13(4): e0193792. <https://doi.org/10.1371/journal.pone.0193792>

Editor: Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: November 13, 2017

Accepted: February 16, 2018

Published: April 18, 2018

Copyright: © 2018 Sun et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Microarray data used in these analyses is available through the Gene Expression Omnibus (accession number GSE111601). All raw metabolomics data is publicly available at the NIH Common Fund Metabolomics Data Repository and Coordinating Center website, Metabolomics Workbench: <http://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000054&StudyType=MS&ResultType=1>.

Abstract

Breast carcinogenesis is a multistep process accompanied by widespread molecular and genomic alterations, both in tumor and in surrounding microenvironment. It is known that tumors have altered metabolism, but the metabolic changes in normal or cancer-adjacent, nonmalignant normal tissues and how these changes relate to alterations in gene expression and histological composition are not well understood. Normal or cancer-adjacent normal breast tissues from 99 women of the Normal Breast Study (NBS) were evaluated. Data of metabolomics, gene expression and histological composition was collected by mass spectrometry, whole genome microarray, and digital image, respectively. Unsupervised clustering analysis determined metabolomics-derived subtypes. Their association with genomic and histological features, as well as other breast cancer risk factors, genomic and histological features were evaluated using logistic regression. Unsupervised clustering of metabolites resulted in two main clusters. The metabolite differences between the two clusters suggested enrichment of pathways involved in lipid metabolism, cell growth and proliferation, and migration. Compared with Cluster 1, subjects in Cluster 2 were more likely to be obese (body mass index ≥ 30 kg/m², $p < 0.05$), have increased adipose proportion ($p < 0.01$) and associated with a previously defined Active genomic subtype ($p < 0.01$). By the integrated analyses of histological, metabolomics and transcriptional data, we characterized two distinct subtypes of non-malignant breast tissue. Further research is needed to validate our findings, and understand the potential role of these alternations in breast cancer initiation, progression and recurrence.

Funding: This research was supported by the Avon Foundation for Women (<https://www.avonfoundation.org/grants/>, PI-MAT), North Carolina University Cancer Research Fund (<http://unclineberger.org/ucrf/>, PI-MAT), National Cancer Institute (R01 CA179715, PI-MAT), the National Institute of Environmental Health Sciences (U01 ES019472, PI-MAT); and NIH Common Fund award through the National Institute for Diabetes and Digestive and Kidney Diseases to the NIH Eastern Regional Comprehensive Metabolomics Resource Core at UNC-Chapel Hill, Nutrition Research Institute (1U24DK097193, PI-SJS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funders provided support in the form of salaries for authors [XS, DAS, RS, ELK, WWP, SLM, RC, MAT and SJS], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: Dr. Robert Clark is employed at the not-for-profit, contract research organization RTI International. His employment and affiliation with this entity did not impact the design, experimental conduct, analysis or decision to publish these results. His affiliation does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

Breast carcinogenesis is associated with extensive changes in the surrounding microenvironment at molecular, genomic and metabolic levels [1–4]. Our previous analysis using gene expression profiling of cancer-adjacent, nonmalignant normal tissue has defined two subtypes, termed Active and Inactive, and showed significant differences in fibrosis, wound response, and cellular adhesion [3]. Subsequent analyses further linked the Active genomic subtype to high adipose content and adverse prognosis [4, 5]. These findings underscore that considering multiple layers of data is an approach for validation, but provides novel insights into the mechanistic networks perturbed during tumorigenesis.

Although extensively studied in breast tumors, metabolic changes in surrounding microenvironment (primarily composed by histologically, nonmalignant normal breast tissue) have been less described, and little is known about how these changes relate to other level changes (e.g. gene expression) and the characteristics of tumor or patient. In this study, we hypothesized that histology, molecular pathways, and metabolic profiles vary, creating distinct niches for the development and progression of breast cancer. To test this hypothesis, we analyzed gene expression, metabolic profiles, and histological composition of normal or cancer-adjacent normal breast tissues from 99 women who participated in the Normal Breast Study (NBS) [6], particularly aimed to characterize metabolic microenvironments and correlates them to other level features.

Materials and methods

Study population

The Normal Breast Study (NBS) was a hospital-based study of breast cancer microenvironment and normal breast tissue, including women undergoing breast surgeries (mastectomy, lumpectomy, excisional biopsy, reduction mammoplasty, or other cosmetic breast surgery) at UNC Hospitals in Chapel Hill, North Carolina from 2009 to 2013 [6]. Fresh non-neoplastic normal breast tissues confirmed by pathology assistants at UNC Hospitals were collected at the time of breast surgery and snap-frozen in liquid nitrogen. Information on clinicopathologic, demographic, and anthropometric factors were collected from medical records and telephone interview. The present analysis included 99 women (16 with reduction mammoplasty, 83 with breast cancer). Written informed consent was obtained from all participants. Study protocols were approved by the School of Medicine Institutional Review Board at the University of North Carolina at Chapel Hill.

Breast tissue processing

Frozen breast specimens of approximately 100 mg were prepared on dry ice and 20 μ m sections were collected at both ends of the specimen and mounted on glass slides. The two sections on the ends were used for histological composition annotation, after hematoxylin and eosin (H&E) staining, and the central tissue portion was used for RNA extraction and whole genome microarray analysis. Additionally, tissues (50–120 mg) were prepared for metabolomics by gas chromatography-time of flight-mass spectrometry (GC-TOF-MS).

Tissue composition measurement

The details regarding histological composition annotation by high-resolution digital images using the Aperio Scan-Scope XT Slide Scanner were published previously [5, 6]. Duplicate slides of same tissue specimen (one section from each end of the sample) were averaged to

estimate area percentage of stroma, adipose and epithelium, given high intra-class correlation coefficients (ICCs) between replicate sections (ICC = ranged from 0.92 = 0.99).

RNA isolation, microarrays, and Active/Inactive subtyping

The central section of fresh-frozen tissues was homogenized using a MagNA Lyser homogenizer (Roche), and RNA was isolated by QIAzol extraction followed by purification on an RNeasy column as described previously [5]. RNA quality and quantity were assessed on an Agilent 2100 Bioanalyzer and a ND-1000 NanoDrop spectrophotometer, respectively. Two-color 4 × 44 K Agilent whole-genome arrays were run on the RNA samples. Data were preprocessed before analytic statistical analyses: lowess-normalization, setting value of the probes that had a signal less than 10 dpi in either channel as missing, excluding probes that had more than 20% missing data across all samples, imputing the missing values using *k*-nearest neighbors' imputation (with *k* = 10), and collapsing the duplicate probes by averaging. Microarray data used in these analyses will be publicly available through the Gene Expression Omnibus (accession number GSE111601) www.ncbi.nlm.nih.gov/geo/.

Samples were classified as Active or Inactive subtype using the Creighton correlation method [5]. Briefly, a standard vector corresponding to all genes (*n* = 3,518) in the Active/Inactive signature was constructed, with "1" assigned to upregulated genes and "-1" assigned to downregulated genes. A Pearson correlation coefficient was calculated for this standard vector versus the vector of median-centered gene expression for each patient. Samples were associated with the Active subtype if the Pearson correlation coefficient was greater than zero; and with the Inactive subtype if the coefficient was less than zero.

Extraction and metabolomics analysis

Tissues were prepared for extractions, metabolomics data collection and preprocessing by the NIH Eastern Regional Comprehensive Metabolomics Resource Core (ERCMRC). The details regarding sample extraction, data collection, and quality control by the ERCMRC have been described previously [7]. Briefly, samples were thawed on ice in a cold room (4°C) and extracted using 50% acetonitrile in water and homogenized on a MagNA Lyser to generate 0.1 mg/μL supernatants. Quality control (QC) samples were created by pooling a small aliquot from all samples (Total Pool 1). A second QC pool sample was created using unequal volumes of samples (Total Pool 2). Samples were chemically derivatized after a D₂O exchange and analyzed by GC-TOF-MS (Agilent 7890 gas chromatogram and Leco Pegasus 4D time of flight mass spectrometer). The run order was randomized and samples were analyzed in four batches over 4 separate days. The Total Pool 1 QC sample was analyzed between every 6 sample injections in each batch. The Total Pool 2 QC sample was injected multiple times at the beginning of each batch to condition the column and the system. The GC-MS data from each sample was normalized to its total signal intensity. Analyte peak variations of the pooled samples were determined, and those peaks with a coefficient of variation (CV) > 0.3 were removed from subsequent analysis. As the data was collected over a four-day period, and preprocessing took account of the day-to-day variability from the GC-TOF-MS by adjusting the peaks from each sample by the median of the value from that of the same peak in the Total Pool 1 samples which were run each day. All raw metabolomics data is publicly available at the NIH Common Fund Metabolomics Data Repository and Coordinating Center website, Metabolomics Workbench: <http://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000054&StudyType=MS&ResultType=1>.

Unsupervised hierarchical clustering analysis (average-linkage) by Cluster 3.0 was used to determine subgroups based on metabolomics profiles. The results were visualized using Java

Treeview. The metabolites significantly different across clusters were identified by supervised multivariate analysis using orthogonal partial least squares discriminant analysis (OPLS-DA) in SIMCA 13.0 (Umetrics, Umeå, Sweden). Metabolites were library-matched to the Fiehn library [8] and rank-ordered based on the value of the Variable Influence on Projection ($VIP \geq 1$). Functional pathway mapping of the known (library-matched) important metabolites was performed using GeneGo (MetaCore-Clarivate Analytics, Philadelphia, PA).

Statistical and pathway analyses

Associations of metabolomics subgroup with age (<40 years, 40-<60 years, or ≥ 60 years), race (white, African American, or others), body mass index (BMI, <30 kg/m² or ≥ 30 kg/m²), tissue type (cancer-adjacent normal tissue or normal tissue), histological adipose composition (<51% or $\geq 51\%$, 51% = median), and Active/Inactive genomic subtype were evaluated qualitatively by Chi-square test or Fisher's exact test, and quantitatively by odds ratio and corresponding 95% confidence interval (CI) estimated using logistic regression. For normal samples from breast cancer patients, the association of metabolomics subgroup with tumor ER status (negative or positive) and the distance to tumor (near, <1cm; or far, >4cm) were also assessed. Probability values of less than 0.05 were considered statistically significant. The analyses were performed using SAS version 9.4 (SAS Institute).

Results

[Table 1](#) describes the characteristics of the NBS participants. Normal breast biospecimens ($n = 83$) from breast cancer patients equaled 84% and 16% ($n = 16$) were from non-breast cancer patients (reduction mammoplasty). The majority of the participants were white (72%, $n = 68$), parous (84%, $n = 80$), and younger than 60 years (72%, $n = 71$). The other 41% ($n = 40$) were premenopausal and 47% ($n = 47$) were obese.

Identification and interpretation metabolomics-derived subgroups

Unsupervised hierarchical clustering of 220 differential metabolites grouped the tissue samples into two metabolomic clusters (Cluster 1, $n = 57$; Cluster 2, $n = 42$) ([Fig 1](#)). No differences were detected in biospecimen type (breast cancer-adjacent normal tissue samples: 84% in Cluster 1, 83% in Cluster 2). Supervised multivariate analysis using SIMCA (OPLS-DA) showed a clear separation of the two clusters based on metabolic profile ([Fig 2A](#)), and identified 102 metabolites based on $VIP \geq 1$, that differentiated Cluster 1 and Cluster 2 ([S1 Table](#)) with 11 known metabolites identified after library-matching to the Fiehn library. The fold-difference of the known metabolites was calculated and three (threonine minor, 2-hydroxybutanoic acid, N-acetyl-L-aspartic acid 1) had a fold change greater than 2, when comparing Cluster 1 to Cluster 2 ([Fig 2B](#)). Enrichment of endogenous pathways were mapped using all 11 library-matched metabolites in GeneGo (MetaCore). [Fig 2C](#) pictorially represents the most significantly enriched pathways, based on the Cluster 1/Cluster 2 fold-change differences and p -value thresholds listed in [S1 Table](#).

Associations with participant characteristics, the Active/Inactive genomic signatures and histological adipose composition

Compared with women in Cluster 1, women in Cluster 2 were older (≥ 60 years, OR = 4.33, 95% CI = 1.14–16.36) and more obese (OR = 2.34, 95% CI = 1.40–5.28) ([Table 1](#)). Tumors of breast cancer patients in Cluster 2 tended to be estrogen receptor (ER) positive (OR = 2.09,

Table 1. Associations with sample characteristics, the Active/Inactive subtype, and histological composition.

	Total n = 99	Cluster 1 n = 57	Cluster 2 n = 42	OR (95% CI)	P-value
Age (year)					
Age ≥ 60	28 (28)	13 (23)	15 (36)	4.33 (1.14, 16.36)	0.08
40 ≤ Age < 60	52 (53)	29 (51)	23 (55)	2.97 (0.87, 10.19)	
Age < 40	19 (19)	15 (26)	4 (10)	1	
Menopausal status					
Premenopausal	40 (41)	25 (44)	15 (37)	0.74 (0.32, 1.68)	0.47
Postmenopausal	58 (59)	32 (56)	26 (63)	1	
Race					
African-American	27 (28)	16 (30)	11 (27)	0.87 (0.35, 2.15)	0.76
White	68 (72)	38 (70)	30 (73)	1	
BMI (kg/m²)					
Obese (BMI ≥ 30)	47 (47)	22 (39)	25 (60)	2.34 (1.40, 5.28)	0.04
Nonobese (BMI < 30)	52 (53)	35 (61)	17 (40)	1	
Parity					
Nulliparous	14 (15)	10 (19)	4 (10)	0.49 (0.14, 1.69)	0.38
Parous	80 (85)	44 (81)	36 (90)	1	
Tissue type					
Cancer Adjacent Normal	83 (84)	48 (84)	35 (83)	0.94 (0.32, 2.76)	0.91
Reduction/Prophylactic	16 (16)	9 (16)	7 (17)	1	
Composition epithelium*					
≥ median	49 (49)	28 (49)	21 (50)	1.04 (0.47, 2.30)	1.0
< median	50 (51)	29 (51)	21 (50)	1	
Composition stroma*					
≥ median	49 (49)	41 (72)	8 (19)	0.09 (0.04, 0.24)	<0.01
< median	50 (51)	16 (28)	34 (81)	1	
Composition adipose*					
≥ median	50 (51)	14 (25)	36 (86)	18.43 (6.42, 52.86)	<0.01
< median	49 (49)	43 (75)	6 (14)	1	
Active/Inactive subtype					
Active	50 (51)	22 (39)	28 (67)	3.18 (1.38, 7.33)	<0.01
Inactive	49 (49)	35 (61)	14 (33)	1	
ER status[†]					
Positive	27 (51)	14 (44)	13 (62)	2.09 (0.68, 6.43)	0.20
Negative	26 (49)	18 (56)	8 (38)	1	
Distance from tumor[†](cm)					
<3	34 (59)	17 (61)	17 (57)	0.85 (0.30, 2.41)	0.75
<1	24 (41)	11 (39)	13 (43)	1	

*Epithelium Median = 7.2% Stroma Median = 36.2% Adipose Median = 50.9%.

[†] Among samples from breast cancer patients. 46 patients without ER information and 41 patients with the distance of normal biospecimen to tumor between 1 and 4 cm were not included the ER and distance analysis respectively.

<https://doi.org/10.1371/journal.pone.0193792.t001>

95% CI = 0.68–6.43). No associations were found between metabolomics subtype and race, menopausal status, parity, or distance to tumor.

Previous research with mRNA and microRNA profiles of these samples demonstrated that there are two main phenotypes of breast tissue, termed Active and Inactive [4]. We evaluated the correlation of these two genomic phenotypes and the metabolomics-derived subtypes, and

found a strong correlation between Cluster 2 and Active subtype (OR = 3.18, 95% CI = 1.38–7.33, $p < 0.01$) (Fig 1 and Table 1).

Obesity and Active/Inactive gene expression subtype are both associated with tissue composition [5], so we hypothesized that adipose content would also vary by metabolomics-derived phenotype. Here, we observed that 86% of the samples in Cluster 2 had high percent adipose content (defined as \geq the median of 51%), compared to 25% of the samples in Cluster 1 (OR = 18.43, 95% CI = 6.42–52.86, $p < 0.01$) (Fig 1 and Table 1). Significant reverse association was observed with stromal contents (OR = 0.09, 95%CI = 0.04–0.24, $p < 0.01$), but not with epithelial content (OR = 1.04, 95%CI = 0.47–2.30, $p = 1.00$).

Discussion

Using unsupervised clustering analysis, we identified two main metabolic clusters in normal breast. Cluster 2, in which the identified metabolites are suggested to play a role in cell death

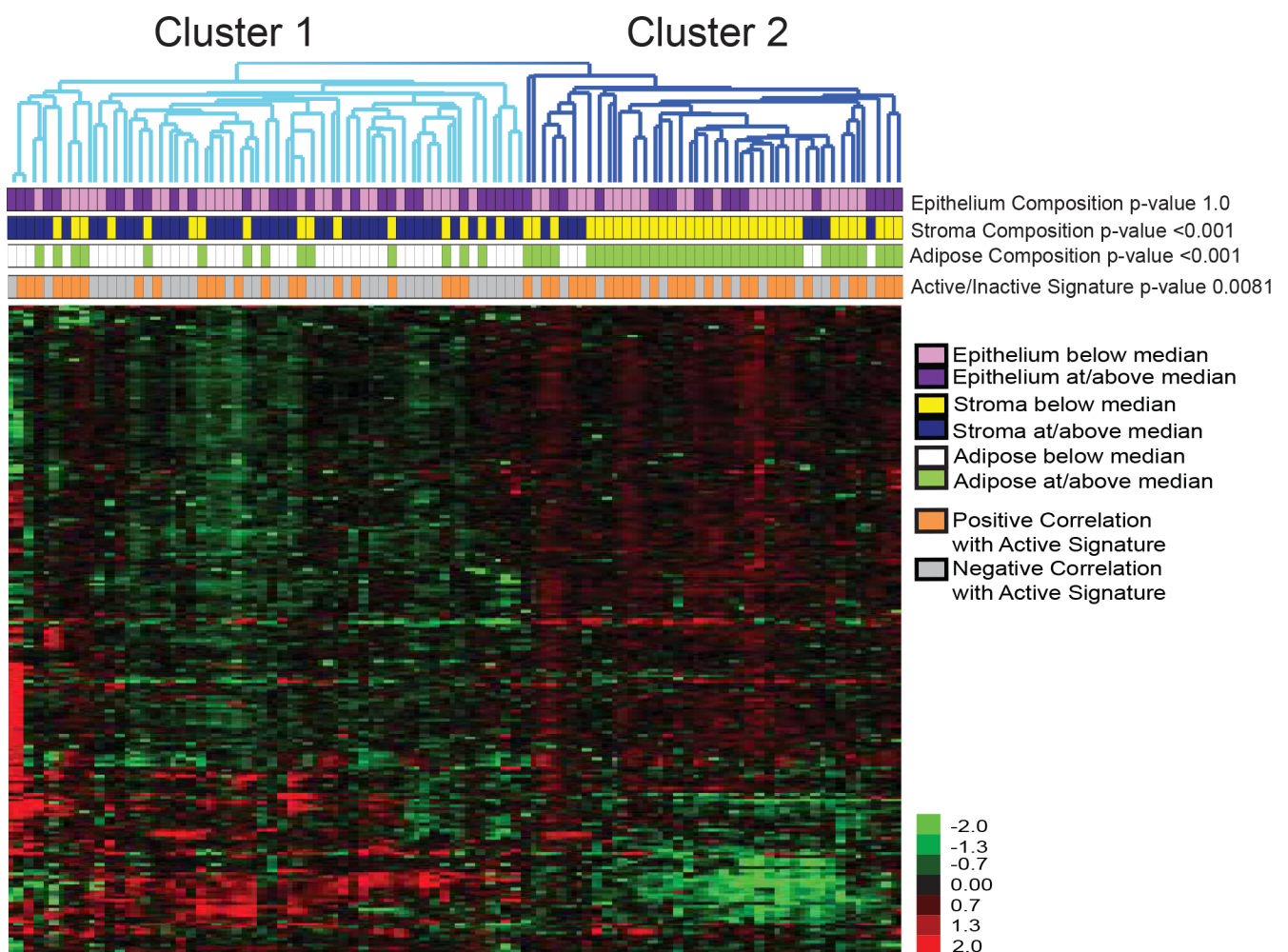


Fig 1. Unsupervised cluster analysis of 220 metabolites across the normal and cancer-adjacent normal breast tissue samples. Pink rectangles represent the samples that have epithelium percent area below the median (7%) for the data set and purple rectangles represent samples that have percent area at or above the median. Yellow rectangles represent the samples that have stroma percent area below the median (36%) for the data set and blue rectangles represent samples that have percent area at or above the median. White rectangles represent the samples that have adipose percent area below the median (51%) for the data set and green rectangles represent samples that have adipose percent area at or above the median for the dataset. Grey rectangles represent the samples that have negative correlation with the active genomic signature and orange rectangles represent samples that have a positive correlation with the active signature.

<https://doi.org/10.1371/journal.pone.0193792.g001>

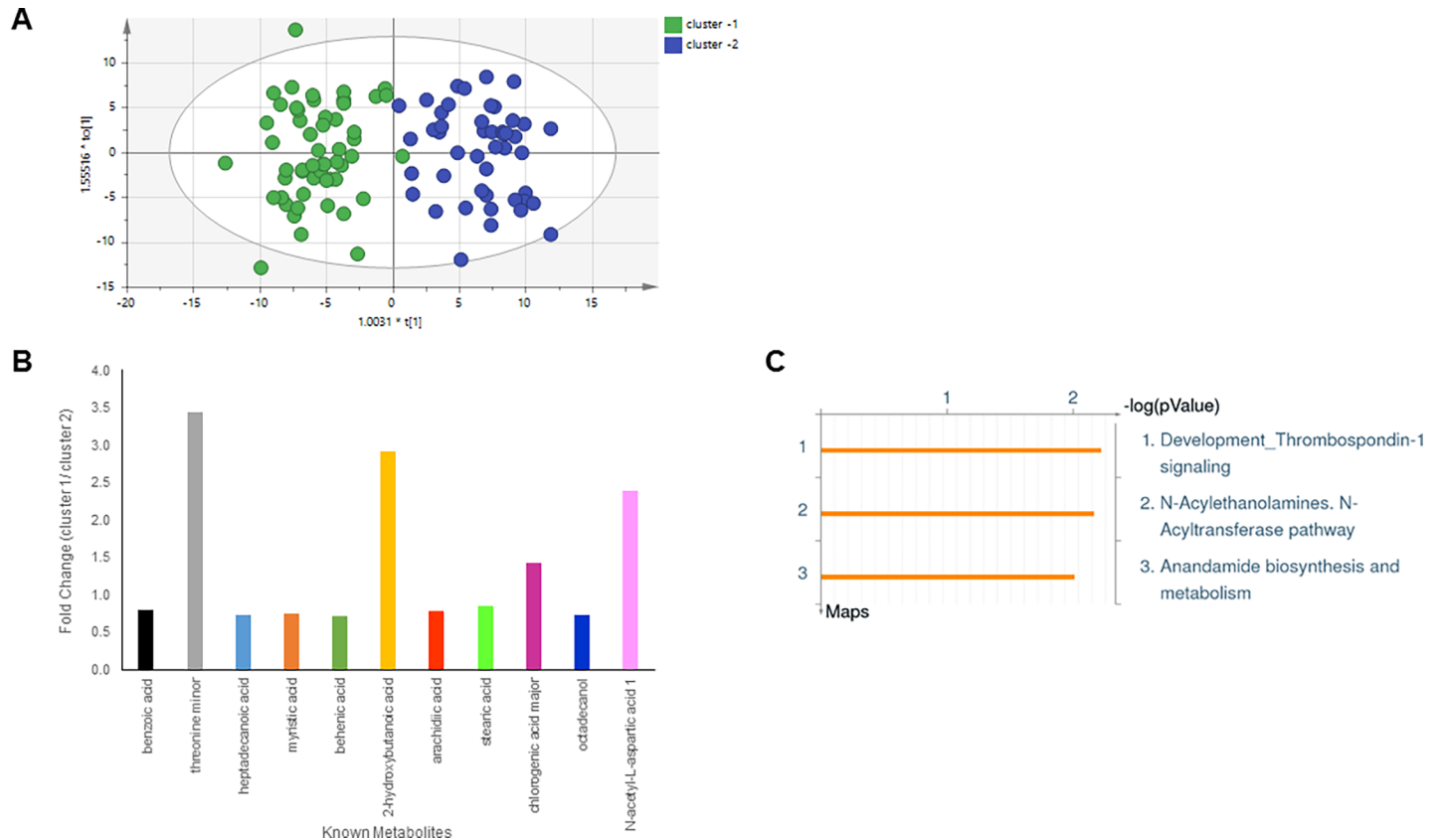


Fig 2. Metabolic Cluster 1 versus Cluster 2. (A) Supervised multivariate analysis (OPLS-DA) of GC-MS generated broad spectrum metabolomics data from non-malignant/cancer-adjacent normal breast tissues. (B) Fold change differences of 11 library-matched metabolites that significantly distinguish metabolic Cluster 1 from Cluster 2. (C) Pathway analysis was done in GeneGo, using the known Cluster 1 –Cluster 2 differentiating metabolites to identify important perturbed biological pathways related to Cluster 1 and Cluster 2 differences.

<https://doi.org/10.1371/journal.pone.0193792.g002>

and molecular transport (Fig 2B), was significantly associated with the Active genomic subtype, as well as obesity and high adipose content in breast tissue.

By using integrated data we demonstrated that multi-layer biological changes are contributing to the breast microenvironment. Brauer et al., previously reported metabolic changes in cancer adjacent normal microenvironment [1]; while in this study, we characterized these changes by defining two distinct metabolic subtypes, and further linked these subtypes to non-malignant genomic breast signatures (Active/Inactive) and tissue adipose content. The Troester lab has previously evaluated the Active/Inactive genomic subtypes for association with mammographic density (MD) and breast tissue composition, and found strong correlations among the Active subtype, high tissue adipose content, and low MD [4, 5]. The current study confirms these previous findings, and extends this correlation circle by adding metabolic Cluster 2 subtype. These consistent findings through integrative data approaches strongly suggest the existence of two broad microenvironmental types of breast tissue. A further question that remains to be addressed is how these microenvironment-defined subtypes mediate an association between risk/prognostic factors, breast cancer occurrence/recurrence and survival outcome, which is very important to understand their role in breast cancer etiology and to identify novel biomarkers to optimize risk prediction and treatment.

Obesity has a dichotomous association with breast cancer, where it has been shown protective in premenopausal women [9], but is an established risk/prognostic factor of breast cancer

in postmenopausal women, with various mechanisms proposed to explain its effect on tumor initiation and progression [10–12]. Our data show that obesity affects the breast histological composition, and is associated with genomic and metabolic changes in histologically normal breast [5, 13], suggesting that obesity profoundly alters the breast microenvironment. In terms of metabolomics profiling, we found that the metabolites higher in Cluster 2 compared to Cluster 1 may be involved in pathways related to lipid metabolism through endocannabinoid-related systems that can modulate cancer cell growth and proliferation (N-Acylethanolamines and Anandamide biosynthesis) [14, 15], and more significantly, creating an environment for increased migratory potential (Thrombospondin 1) [16] (Fig 2C). These findings are in line with previous data, where broad changes in lipid metabolism during breast carcinogenesis have been associated with maintaining cell growth, playing a role in cell survival and contributing to metastasis [17–21]. Our analyses describe the overall correlations of obesity with metabolic and genomic profiles in the normal and cancer-adjacent normal breast microenvironment. However, how these changes at each level interact with one another (e.g. network linkage of specific metabolite(s) to certain gene(s)) needs further study.

Age is the strongest risk factor for breast cancer, and is associated with genomic and histological alternations [6, 22–24]. In this study we observed a significant association between metabolic subtype, when comparing young women (<40 years) and old women (≥ 60 years) (OR = 4.33, 95% CI = 1.14–16.36). This finding is consistent with the data that age is positively associated with the Active genomic signature and high adipose content. However, given this correlation between age and obesity and the limited sample size of our study, these data will not allow us to test whether their effect on microenvironment is independent, and if not, how age and obesity interact across microenvironmental changes.

Several studies have evaluated the association between tumor characteristics (such as molecular subtypes and ER status) and metabolomics profiles of breast tissues. Some investigators observed a separation between ER positive and ER negative tumors based on metabolomics profiles [7, 17, 25], whereas others showed overlapping patterns between ER positive and ER negative samples [1, 26]. In the current study, we observed that metabolic Cluster 2 included more women with ER positive tumors (OR = 2.09). However, our findings should be interpreted with caution in that it was based on univariate analysis (not adjusting for potential confounders, e.g. obesity), and not statistically significant ($p = 0.20$).

Our study should be considered in light of limitations and strengths. First, the sample size of our study may lead to imprecise effect estimates (e.g. OR of adipose composition) and lack of statistical power for multivariate analyses or stratified analyses. Future studies with larger sample size are needed to overcome these caveats and validate our findings. Second, although we detected over a hundred significant compound ions, only 11 could be library-matched to known metabolites for use in pathway analysis. Therefore, the results of pathway analysis do not characterize the two metabolic Clusters comprehensively. One important strength of our study is the use of normal and cancer-adjacent normal breast tissue. Most metabolic studies are done in the diseased tissue/tumor where the characteristics of the pathologic tissue are compared to normal tissue in order to understand the variations from normalcy that accounts for the abnormality manifested as disease/pathology [26, 27]. However, the corresponding changes in normal versus cancer-adjacent normal tissue are not well understood. In our study, we used both normal and cancer-adjacent normal breast tissue samples from the Normal Breast Study (NBS), and provided significant associations between genomic, histologic, and metabolic profiles within a population of women with histologically normal-appearing breast tissue. Our results help to increase understanding on how normal tissue responds to breast cancer risk factor exposures, the adjacent tumor microenvironment, and how these dynamics may mediate breast cancer risk and outcome.

To identify targetable pathways to disrupt the process of tumor growth and progression, future research with sufficient sample size of normal breast tissue is needed to better characterize genomic, metabolomic and histological microenvironments, and to understand how these associated risk factors (e.g. obesity) drive breast tissue microenvironment alterations and ultimately promote disease. In addition, the identification and analysis of additional metabolites that may have a role in modulating the effect of various risk factors will help to further our understanding of normal and cancer-adjacent normal breast phenotypes.

Supporting information

S1 Table. Significant Cluster 1 vs. Cluster 2 metabolites.
(DOCX)

Acknowledgments

This research was supported by the Avon Foundation for Women (<https://www.avonfoundation.org/grants/>, PI-MAT), North Carolina University Cancer Research Fund (<http://unclineberger.org/ucrf/>, PI-MAT), National Cancer Institute (R01 CA179715, PI-MAT), the National Institute of Environmental Health Sciences (U01 ES019472, PI-MAT); and NIH Common Fund award through the National Institute for Diabetes and Digestive and Kidney Diseases to the NIH Eastern Regional Comprehensive Metabolomics Resource Core at UNC-Chapel Hill, Nutrition Research Institute (1U24DK097193, PI-SJS).

Author Contributions

Conceptualization: Rupninder Sandhu, Melissa A. Troester, Susan J. Sumner.

Data curation: Delisha A. Stewart, Rupninder Sandhu, Erin L. Kirk.

Formal analysis: Xuezheng Sun, Delisha A. Stewart, Rupninder Sandhu, Erin L. Kirk, Wimal W. Pathmasiri, Susan L. McRitchie, Robert F. Clark.

Funding acquisition: Melissa A. Troester, Susan J. Sumner.

Investigation: Delisha A. Stewart.

Methodology: Delisha A. Stewart, Wimal W. Pathmasiri, Melissa A. Troester, Susan J. Sumner.

Project administration: Delisha A. Stewart, Susan J. Sumner.

Resources: Melissa A. Troester, Susan J. Sumner.

Validation: Rupninder Sandhu, Susan L. McRitchie.

Visualization: Rupninder Sandhu.

Writing – original draft: Xuezheng Sun, Delisha A. Stewart, Rupninder Sandhu, Robert F. Clark.

Writing – review & editing: Xuezheng Sun, Delisha A. Stewart, Rupninder Sandhu, Wimal W. Pathmasiri, Susan L. McRitchie, Robert F. Clark, Melissa A. Troester, Susan J. Sumner.

References

1. Brauer HA, Makowski L, Hoadley KA, Casbas-Hernandez P, Lang LJ, Roman-Perez E, et al. Impact of tumor microenvironment and epithelial phenotypes on metabolism in breast cancer. *Clinical cancer*

- research: an official journal of the American Association for Cancer Research. 2013; 19(3):571–85. <https://doi.org/10.1158/1078-0432.CCR-12-2123> PMID: 23236214
2. Casbas-Hernandez P, Sun X, Roman-Perez E, D'Arcy M, Sandhu R, Hishida A, et al. Tumor intrinsic subtype is reflected in cancer-adjacent tissue. *Cancer Epidemiol Biomarkers Prev.* 2015; 24(2):406–14. Epub 2014/12/04. <https://doi.org/10.1158/1055-9965.EPI-14-0934> PMID: 25465802; PubMed Central PMCID: PMC4437571.
 3. Roman-Perez E, Casbas-Hernandez P, Pirone JR, Rein J, Carey LA, Lubet RA, et al. Gene expression in extratumoral microenvironment predicts clinical outcome in breast cancer patients. *Breast cancer research: BCR.* 2012; 14(2):R51. <https://doi.org/10.1186/bcr3152> PMID: 22429463
 4. Troester MA, Hoadley KA, D'Arcy M, Cherniack AD, Stewart C, Koboldt DC, et al. DNA defects, epigenetics, and gene expression in cancer-adjacent breast: a study from The Cancer Genome Atlas. *Npj Breast Cancer.* 2016; 2:16007. <https://doi.org/10.1038/npjbcancer.2016.7> PMID: 28721375
 5. Sun X, Gierach GL, Sandhu R, Williams T, Midkiff BR, Lissowska J, et al. Relationship of mammographic density and gene expression: analysis of normal breast tissue surrounding breast cancer. *Clin Cancer Res.* 2013; 19(18):4972–82. Epub 2013/08/07. <https://doi.org/10.1158/1078-0432.CCR-13-0029> PMID: 23918601; PubMed Central PMCID: PMC4073678.
 6. Sandhu R, Chollet-Hinton L, Kirk EL, Midkiff B, Troester MA. Digital histologic analysis reveals morphometric patterns of age-related involution in breast epithelium and stroma. *Human pathology.* 2016; 48:60–8. Epub 2016/01/17. <https://doi.org/10.1016/j.humpath.2015.09.031> PMID: 26772400; PubMed Central PMCID: PMC4727830.
 7. Budczies J, Brockmoller SF, Muller BM, Barupal DK, Richter-Ehrenstein C, Kleine-Tebbe A, et al. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *Journal of proteomics.* 2013; 94:279–88. Epub 2013/10/16. <https://doi.org/10.1016/j.jprot.2013.10.002> PMID: 24125731.
 8. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, et al. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical chemistry.* 2009; 81(24):10038–48. Epub 2009/11/26. <https://doi.org/10.1021/ac9019522> PMID: 19928838; PubMed Central PMCID: PMC4073678.
 9. Research WCRFAfC. Continuous Update Project Report: Diet, Nutrition, Physical Activity and Breast Cancer. wcrf.org/breast-cancer-sl-2017 2017.
 10. Sun X, Nichols HB, Robinson W, Sherman ME, Olshan AF, Troester MA. Post-diagnosis adiposity and survival among breast cancer patients: influence of breast cancer subtype. *Cancer causes & control: CCC.* 2015; 26(12):1803–11. Epub 2015/10/03. <https://doi.org/10.1007/s10552-015-0673-6> PMID: 26428518; PubMed Central PMCID: PMC4073678.
 11. Sinicropo FA, Dannenberg AJ. Obesity and breast cancer prognosis: weight of the evidence. *J Clin Oncol.* 2011; 29(1):4–7. Epub 2010/12/01. <https://doi.org/10.1200/JCO.2010.32.1752> PMID: 21115867.
 12. Jiralerspong S, Goodwin PJ. Obesity and Breast Cancer Prognosis: Evidence, Challenges, and Opportunities. *J Clin Oncol.* 2016; 34(35):4203–16. Epub 2016/12/03. <https://doi.org/10.1200/JCO.2016.68.4480> PMID: 27903149.
 13. Sun X, Casbas-Hernandez P, Bigelow C, Makowski L, Joseph Jerry D, Smith Schneider S, et al. Normal breast tissue of obese women is enriched for macrophage markers and macrophage-associated gene expression. *Breast cancer research and treatment.* 2012; 131(3):1003–12. Epub 2011/10/18. <https://doi.org/10.1007/s10549-011-1789-3> PMID: 22002519; PubMed Central PMCID: PMC4073678.
 14. Brown I, Wahle KW, Cascio MG, Smoum-Jaouni R, Mechoulam R, Pertwee RG, et al. Omega-3 N-acyl-ethanolamines are endogenously synthesised from omega-3 fatty acids in different human prostate and breast cancer cell lines. *Prostaglandins Leukot Essent Fatty Acids.* 2011; 85(6):305–10. Epub 2011/10/15. <https://doi.org/10.1016/j.plefa.2011.09.007> PMID: 21995886.
 15. Gaston R, Maria Eugenia P, Das UN, Eynard AR. Polyunsaturated Fatty Acids Differentially Modulate Cell Proliferation and Endocannabinoid System in Two Human Cancer Lines. *Arch Med Res.* 2017; 48(1):46–54. Epub 2017/06/05. <https://doi.org/10.1016/j.arcmed.2017.01.009> PMID: 28577869.
 16. Chandrasekaran S, Guo NH, Rodrigues RG, Kaiser J, Roberts DD. Pro-adhesive and chemotactic activities of thrombospondin-1 for breast carcinoma cells are mediated by alpha3beta1 integrin and regulated by insulin-like growth factor-1 and CD98. *J Biol Chem.* 1999; 274(16):11408–16. Epub 1999/04/10. PMID: 10196234.
 17. Hilvo M, Denkert C, Lehtinen L, Muller B, Brockmoller S, Seppanen-Laakso T, et al. Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res.* 2011; 71(9):3236–45. Epub 2011/03/19. <https://doi.org/10.1158/0008-5472.CAN-10-3894> PMID: 21415164.

18. Louie SM, Roberts LS, Mulvihill MM, Luo K, Nomura DK. Cancer cells incorporate and remodel exogenous palmitate into structural and oncogenic signaling lipids. *Biochim Biophys Acta*. 2013; 1831(10):1566–72. Epub 2013/07/23. <https://doi.org/10.1016/j.bbaliip.2013.07.008> PMID: 23872477; PubMed Central PMCID: PMC3838875.
19. Kuemmerle NB, Rysman E, Lombardo PS, Flanagan AJ, Lipe BC, Wells WA, et al. Lipoprotein lipase links dietary fat to solid tumor cell proliferation. *Molecular cancer therapeutics*. 2011; 10(3):427–36. Epub 2011/02/02. <https://doi.org/10.1158/1535-7163.MCT-10-0802> PMID: 21282354; PubMed Central PMCID: PMC3074101.
20. Nomura DK, Long JZ, Niessen S, Hoover HS, Ng SW, Cravatt BF. Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell*. 2010; 140(1):49–61. Epub 2010/01/19. <https://doi.org/10.1016/j.cell.2009.11.027> PMID: 20079333; PubMed Central PMCID: PMC2885975.
21. Mishra P, Ambs S. Metabolic Signatures of Human Breast Cancer. *Molecular & cellular oncology*. 2015; 2(3). Epub 2015/05/26. <https://doi.org/10.4161/23723556.2014.992217> PMID: 26005711; PubMed Central PMCID: PMC34438683.
22. Nassar A, Visscher DW, Degnim AC, Frank RD, Vierkant RA, Frost M, et al. Complex fibroadenoma and breast cancer risk: a Mayo Clinic Benign Breast Disease Cohort Study. *Breast cancer research and treatment*. 2015; 153(2):397–405. Epub 2015/08/13. <https://doi.org/10.1007/s10549-015-3535-8> PMID: 26264469; PubMed Central PMCID: PMC34561026.
23. Figueroa JD, Pfeiffer RM, Patel DA, Linville L, Brinton LA, Gierach GL, et al. Terminal duct lobular unit involution of the normal breast: implications for breast cancer etiology. *Journal of the National Cancer Institute*. 2014; 106(10). Epub 2014/10/03. <https://doi.org/10.1093/jnci/dju286> PMID: 25274491; PubMed Central PMCID: PMC34200067.
24. Pirone JR, D'Arcy M, Stewart DA, Hines WC, Johnson M, Gould MN, et al. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2012; 21(10):1735–44. Epub 2012/08/04. <https://doi.org/10.1158/1055-9965.EPI-12-0451> PMID: 22859400; PubMed Central PMCID: PMC3684707.
25. Terunuma A, Putluri N, Mishra P, Mathe EA, Dorsey TH, Yi M, et al. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J Clin Invest*. 2014; 124(1):398–412. Epub 2013/12/10. <https://doi.org/10.1172/JCI71180> PMID: 24316975; PubMed Central PMCID: PMC3871244.
26. Tang X, Lin CC, Spasojevic I, Iversen ES, Chi JT, Marks JR. A joint analysis of metabolomics and genetics of breast cancer. *Breast Cancer Res*. 2014; 16(4):415. Epub 2014/08/06. <https://doi.org/10.1186/s13058-014-0415-9> PMID: 25091696; PubMed Central PMCID: PMC34187326.
27. Penkert J, Ripperger T, Schieck M, Schlegelberger B, Steinemann D, Illig T. On metabolic reprogramming and tumor biology: A comprehensive survey of metabolism in breast cancer. *Oncotarget*. 2016; 7(41):67626–49. Epub 2016/09/04. <https://doi.org/10.18632/oncotarget.11759> PMID: 27590516; PubMed Central PMCID: PMC35341901.