Human Pathology (2022) **123**, 84–92

ELSEVIER

Human PATHOLOGY

www.elsevier.com/locate/humpath

Original contribution

Quantitative analysis of breast cancer tissue composition and associations with tumor subtype☆,☆☆

Linnea T. Olsson MSPH, BS, BA [a], Lindsay A. Williams MPH, PhD [b,c], Bentley R. Midkiff BS [d], Erin L. Kirk MS [a], Melissa A. Troester PhD, MPH [a,e,1], Benjamin C. Calhoun MD, PhD [e,f,*,1]

[a] *Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 27599, USA*
[b] *Department of Pediatrics, University of Minnesota School of Medicine, Minneapolis, MN, 55455, USA*
[c] *Masonic Cancer Center, University of Minnesota, Minneapolis, MN, 55455, USA*
[d] *Pathology Services Core, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599, USA*
[e] *Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599, USA*
[f] *Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC, 27599, USA*

Received 4 January 2022; revised 14 February 2022; accepted 16 February 2022

Available online 23 February 2022

Keywords:
Breast cancer;
Microenvironment;
Pathology;
Digital histology;
Stroma;
Tumor-infiltrating lymphocytes (TILs)

Summary The tumor microenvironment is an important determinant of breast cancer progression, but standard methods for describing the tumor microenvironment are lacking. Measures of microenvironment composition such as stromal area and immune infiltrate are labor-intensive and few large studies have systematically collected this data. However, digital histologic approaches are becoming more widely available, allowing high-throughput, quantitative estimation. We applied such methods to tissue microarrays of tumors from 1687 women (mean 4 cores per case) in the Carolina Breast Cancer Study Phase 3. Tumor composition was quantified as percentage of epithelium, stroma, adipose, and lymphocytic infiltrate (with the latter as presence/absence using a $\geq$1% cutoff). Composition proportions and presence/absence were evaluated in association with clinical and molecular features of breast cancer

☆ Competing interests: Dr. Calhoun is a member of the Oncology Advisory Board for Luminex Corp. All other authors declare that there are no conflicts of interest to report. The University of North Carolina, Chapel Hill has a license of intellectual property interest in GeneCentric Diagnostics and BioClassifier, LLC, which may be used in this study. The University of North Carolina, Chapel Hill may benefit from this interest that is/are related to this research. The terms of this arrangement have been reviewed and approved by the University of North Carolina, Chapel Hill Conflict of Interest Program in accordance with its conflict of interest policies.

☆☆ Funding/Support: This work was supported by P30 ES010126, U01 CA179715, P50 CA058223, U01 ES19472, R01 CA253450, and the Susan G. Komen for the Cure, including a Komen Foundation Graduate Training in Disparities Research program grant.

* Corresponding author. 160 N. Medical Drive, CB #7435, Chapel Hill, NC, 27599, USA.
E-mail address: ben.calhoun@unchealth.unc.edu (B.C. Calhoun).

[1] Co-senior authors.

https://doi.org/10.1016/j.humpath.2022.02.013

(intrinsic subtype and RNA-based risk of recurrence [ROR] scores) using multivariable linear and logistic regression. Lower stromal content was associated with aggressive tumor phenotypes, including triple-negative (31.1% vs. 41.6% in HR+/HER2-; RFD [95% CI]: −10.5%, [-13.1, −7.9]), Basal-like subtypes (29.0% vs. 44.0% in Luminal A; RFD [95% CI]: −14.9%, [-17.8, −12.0]), and high RNA-based PAM50 ROR scores (27.6% vs. 48.1% in ROR low; RFD [95% CI]: −20.5%, [24.3, 16.7]), after adjusting for age and race. HER2+ tumors also had lower stromal content, particularly among RNA-based HER2-enriched (35.2% vs. 44.0% in Luminal A; RFD [95% CI]: −8.8%, [-13.8, −3.8]). Similar associations were observed between immune infiltrate and tumor phenotypes. Quantitative digital image analysis of the breast cancer microenvironment showed significant associations with demographic characteristics and biological indicators of aggressive behavior.

1. Introduction

Classification schema for breast tumors most commonly emphasize molecular markers detected in epithelium, and until recently, non-epithelial cancer-adjacent tissues were not considered as primary drivers of tumorigenesis and cancer progression. However, the importance of the immune microenvironment is now widely recognized, and it is increasingly evident that other non-immune stromal components in and around the tumor may also play a role in cancer development and progression [2−7]. Recent studies have observed associations between stroma- and immune-associated gene expression in tumor samples and clinical outcomes among breast cancer patients [8−10]. Stroma interacts with different breast cancer subtypes in characteristic ways in cell-based assays [22−24], suggesting that cells in the stroma play an understudied role in tumor biology. Simultaneously, rapidly growing tumor cells may also rapidly overtake stroma within a tumor, so the empirical observation of absence of intratumoral stroma may itself be a poor prognostic feature. With respect to immune cells, triple negative and HER2+ subtypes have been shown to have higher immune infiltration by lymphocytes, a pattern which may predict response to chemotherapy [16]. However, few large, population-based studies have been able to systematically evaluate these two important microenvironment features in cancers. This is in part because visual estimation of stromal and immune components is time consuming and largely qualitative.

Application of digital histology methods with curated features [1,2] has allowed for more facile, quantitative characterization of tumor composition. We applied quantitative digital histology methods to tissue microarrays (TMA) from the Carolina Breast Cancer Study Phase 3 [3,4], a study that oversampled Black and young women who are likely to have more aggressive breast cancer. We hypothesized that the proportion of both stroma and immune infiltrate in the tumor microenvironment may be associated with clinical and molecular markers of breast cancer prognosis.

2. Methods

2.1. Study population

The Carolina Breast Cancer Study (CBCS3) Phase 3 is a population-based, prospective cohort of women with incident breast cancer in 44 counties of North Carolina. Details on the recruitment, data collection procedures and molecular characterization of the tumors in CBCS3 have been described [3,4]. Briefly, cases were identified using a rapid case ascertainment system developed in collaboration with the North Carolina Central Cancer Registry. Eligible participants included women between the ages of 20 and 74 diagnosed with a first primary invasive breast cancer from May 1, 2008 to July 31, 2013. Black and young (<50 years old) women were oversampled to assess heterogeneity of disease according to population characteristics and to identify determinants of breast cancer outcome disparities. The study was approved by the University of North Carolina (UNC) School of Medicine Institutional Review Board and participants provided written informed consent (IRB number 92−0410).

2.2. Tissue microarrays

Formalin-fixed, paraffin-embedded (FFPE) tumor blocks were obtained from participating medical centers for participants if sufficient tumor tissue remained after clinical management. Of the 2998 women enrolled in the study, tumor tissue was obtained from 95% of participants, including either slides or blocks. For TMA construction in this analysis, blocks were emphasized, and blocks were obtained for 1743 women. Up to three tumor blocks per woman were obtained, with the representative block selected by the study pathologist. Most women (approximately 95%) had only one block available. Central review by a board-certified pathologist identified and annotated tumor-enriched areas for construction of TMAs. From these regions, between one and four 1-mm tumor cores were taken to be embedded in TMAs for immunohistochemistry and image analysis. Cores were extracted from tumor

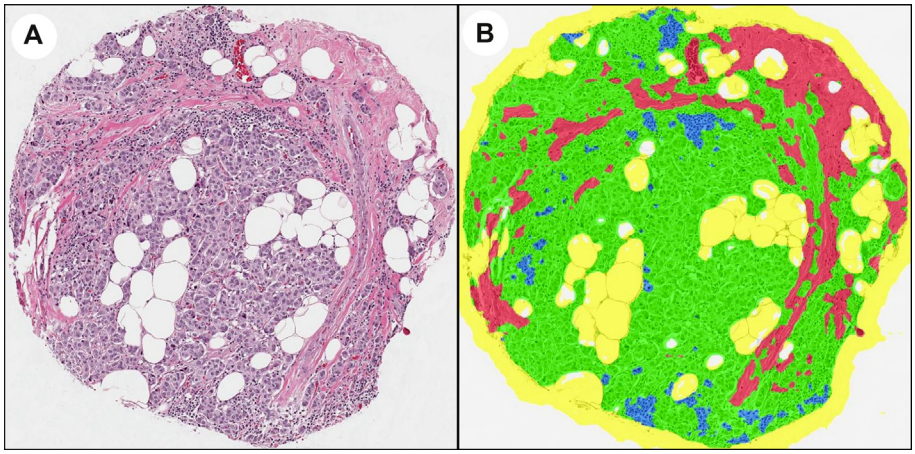

Fig. 1 A representative core section (A) without and (B) with the Genie tissue composition algorithm overlaid. All four components – epithelium (green), stroma (red), adipose (yellow), and immune infiltrate (blue) – are visible on this slide. An excess ring of adipose classification is present which will be analytically subtracted from the analysis area (see Fig. 2). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

blocks and embedded into TMAs by the UNC-Chapel Hill Translational Pathology Laboratory (TPL). The TMA blocks were serially cut into 4 μm-thick sections and mounted on slides. The top and bottom sections were stained with hematoxylin and eosin (H&E) to assess tumor cellularity at both depths. Sections were then scanned at 20× magnification into the Aperio eSlide Manager (Leica Biosystems) digital slide database at TPL using the Aperio Scan-Scope XT Slide Scanner (Leica Biosystems).

Quality control (QC) was conducted prior to image analysis to exclude cores that were of poor quality or that represented mucinous tumors (n = 55) that were not amenable to digital analysis. Poor quality cores included those with degraded tissue and sections in which red blood cells, cysts, or necrosis made up a large proportion (>50%) of the core. Mucinous tumors could not be distinguished from adipose tissue, so all cores from those patients were excluded. When low quality cores were identified for other reasons, only individual cores were excluded and other cores that passed QC were retained in the dataset. After core exclusions, a total of 1687 women remained in the study population, with a median of 8 core sections per participant.

2.3. Training and testing of the genie algorithm on CBCS3

An initial set of 10 TMA slides was selected to train a tissue composition classifier. First, these slides were manually annotated in Aperio ImageScope (Leica Biosystems) with discrete areas of representative of epithelium, stroma, adipose, and lymphocytic immune infiltrate. The Aperio Genie algorithm (Leica Biosystems) was provided with these training slides as a feature library, which was then used to build the classifier [1]. Modifications were made to adjustment parameters such as maximum or minimum cell size and stain intensity to improve the algorithm. Visual inspection of tissue segmentation was evaluated with each parameter adjustment. Finally, the classifier was applied to all 176 TMA slides (corresponding to the top and bottom sections from 88 TMA blocks). A second algorithm was developed using similar methodology in the Definiens Tissue Studio Software (Biocompare) for a subset of 10 TMAs to evaluate consistency across available tissue segmentation platforms. The second algorithm developed using Definiens was strongly correlated with the first (Aperio), with correlations of 0.82 (epithelium), 0.84 (stroma), and 0.83 (adipose). The two classifiers had 73% agreement for presence/absence of immune infiltrate. Both algorithms showed some low frequency misclassification of tumor cells with pyknotic nuclei admixed with lymphocytes (Fig. 1). The study pathologist interpreted differences between the algorithms as suggestive of random error and did not find evidence that either algorithm was biased with respect to lymphocyte quantification.

To assign histologic composition (area of each tissue type), analysis was performed separately for every core (both a top H&E and a bottom H&E, constructed after sectioning of 10 unstained slides), and individual estimates were computed as an area-weighted average over all core images available for the case. First, glass was excluded from the analysis area for each core. Then, total tissue area was calculated across all cores for each case. The individual tissue components were summed across all cores for each case, and the percentage of each component was calculated as the ratio of the area of the tissue component to the total tissue area times 100. We noted that each core had a small circumferential area (radial distance of 0.0403 mm) that was misclassified as adipose (rather than glass), and this was corrected in final estimates of adipose and total area by subtraction.

2.4. Statistical analysis

Two histological parameters − percent of area of core comprised of stroma (continuous) and immune infiltrate (dichotomized as present (≥1%) or absent (<1%)) − were used as dependent variables in this analysis. Coefficients of variation (CV) were estimated for each tissue class to characterize the magnitude of variability. Multivariable linear models with a Gaussian error distribution were fit to estimate mean differences and 95% confidence intervals for the association between percent stroma (as a continuous variable) and tumor characteristics. Multivariable generalized linear regression models with a binomial error distribution and identity link were constructed to estimate relative frequency differences (RFD) and 95% confidence intervals for the association between lymphocytic immune infiltrate (binary variable) and tumor characteristics.

Tumor variables included IHC subtype (HR+/HER2-, HER2+, triple-negative), PAM50 subtype (Luminal A, Luminal B, HER2-enriched, Basal-like), ROR-PT score (low/intermediate/high), combined histologic grade (low/intermediate/high), and histologic subtype (ductal/lobular). After Bonferroni-Holm adjustment for multiple testing, p-values were two-sided with an alpha value of 0.05 as a threshold for statistical significance. All models were adjusted for binary age (<50, ≥50 years old) and race (Black/non-Black) variables.

Finally, as an exploratory analysis, we used a receiver operating characteristic curve to illustrate the potential ability of stromal proportion to distinguish Basal-like from non-Basal-like tumors and, among ER-positive tumors, to distinguish Luminal A from Luminal B tumors. ROC curve analysis was performed in R using the pROC package, and area under the curve (AUC) was calculated.

3. Results

3.1. Study population and digital image analysis

The clinical, demographic and histological data for the study population are summarized in Table 1. Overall, the study population had a mean age of 52.7 and approximately equal proportions of Black (n = 811) and non-Black (n = 876) participants. The majority of tumors were early stage, intermediate and high-grade, and hormone-receptor-positive (HR+) and HER2-negative by immunohistochemistry (IHC). The subset of women with available images did not differ from the full CBCS3 population.

A representative core section with each of the four histologic classes present is shown in Fig. 1A, while Fig. 1B presents the same core section with the selected Genie tissue composition algorithm overlaid. Overall, epithelium was the most highly represented (and least variable) histologic class, with a mean proportion of almost 50% in the population, followed by stroma (34%), adipose tissue (13%), and finally immune infiltrate (3%). The coefficient of variation was highest for immune infiltrate (CV = 2.20) and lowest for epithelium (CV = 0.46).

Table 1 Histological, demographic, and clinical characteristics of the study population (n = 1687) and full CBCS3 population (n = 2998).

	Study Population N (%) or Mean (SD)	Full CBCS3 Population N (%) or Mean (SD)
Stroma (continuous)	34.6 (21.3)	
Lymphocytic immune infiltrate		
<1%	953 (56.5)	
≥1%	734 (43.5)	
Age (continuous)	52.7 (11.1)	51.7 (11.1)
Race		
Black	811 (48.1)	1495 (49.9)
Non-Black	876 (51.9)	1503 (50.1)
Stage at diagnosis		
I	760 (44.7)	1254 (41.9)
II	687 (41.0)	1205 (40.2)
III	199 (11.9)	427 (14.3)
IV	39 (2.3)	109 (3.6)
Missing	0	3
IHC Subtype		
HR + HER2-	1114 (67.3)	1901 (65.0)
HER2+	223 (13.5)	441 (15.1)
Triple-negative	317 (19.2)	583 (19.9)
Missing	33	73
PAM50 Subtype		
Luminal A	608 (52.1)	692 (47.7)
Luminal B	195 (16.7)	236 (16.3)
HER2-enriched	66 (5.7)	120 (8.3)
Basal-like	275 (23.6)	373 (25.7)
Normal-like	22 (1.9)	29 (2.0)
Missing	521	1548
Risk of Recurrence Score		
Low	217 (18.9)	266 (18.7)
Intermediate	700 (60.8)	842 (59.3)
High	234 (20.3)	313 (22.0)
Missing	536	1577
Combined Histologic Grade		
Low	333 (19.9)	604 (20.4)
Intermediate	628 (37.5)	1138 (38.4)
High	713 (42.6)	1224 (41.3)
Missing	13	32
Histological Subtype		
Ductal	1496 (89.3)	2661 (89.4)
Lobular	180 (10.7)	315 (10.6)
Missing	11	22

3.2. Associations between tissue composition and tumor characteristics

3.2.1. Stromal composition

Continuous measures of stroma proportion were inversely associated with aggressive disease phenotypes (Table 2,

Table 2 Associations between continuous measures of stromal proportion and tumor and clinical characteristics of the study population.

	Stromal Proportion (95% CI)[a]	Percent Difference (95% CI)[a]	*p*-value[b]
IHC Subtype			
HR + HER2-	41.6 (39.9, 43.4)	REF	REF
HER2+	38.5 (35.4, 41.6)	−3.1 (−6.1, −0.2)	0.0391
Triple-Negative	31.1 (28.3, 33.9)	−10.5 (−13.1, −7.9)	<0.0001
PAM50 Subtype[c]			
Luminal A	44.0 (41.9, 46.1)	REF	REF
Luminal B	33.5 (29.1, 35.8)	−11.5 (−14.7, −8.3)	<0.0001
HER2-enriched	35.2 (30.1, 40.3)	−8.8 (−13.8, −3.8)	0.0013
Basal-like	29.0 (26.0, 32.1)	−14.9 (−17.8, −12.0)	<0.0001
ROR-PT Score			
Low	48.1 (45.2, 51.0)	REF	REF
Intermediate	37.8 (35.6, 40.0)	−10.3 (−13.4, −7.3)	<0.0001
High	27.6 (24.2, 31.0)	−20.5 (−24.3, −16.7)	<0.0001
Grade			
Low	47.7 (45.3, 50.0)	REF	REF
Intermediate	42.2 (40.3, 44.1)	−5.5 (−8.1, −2.9)	0.0001
High	28.4 (26.2, 30.5)	−19.3 (−21.9, −16.7)	<0.0001
Morphology			
Ductal	32.8 (31.7, 33.8)	REF	
Lobular	49.0 (46.0, 52.0)	16.2 (13.0, 19.4)	<0.0001

[a] Models adjusted for age and race ($p < 0.05$).
[b] *P*-values adjusted for multiple testing using Bonferroni-Holm method.
[c] Normal-like samples were excluded for the purposes of PAM50 assessment because of small sample size.

Fig. 2A). Clinically triple-negative (i.e., HR-negative/HER2-negative by IHC) tumors had the lowest proportion of stroma (31.1%) of all subtypes. The highest stromal proportion was seen in hormone receptor (HR)-positive, HER2-negative tumors (41.6%). HER2-positive tumors had a mean stroma proportion intermediate between HR-positive, HER2-negative and triple-negative tumors [38.5%; mean difference = −3.1 (95% CI: −6.1, −0.2)]. Similar results were obtained with PAM50-derived intrinsic molecular subtypes. Basal-like tumors had significantly lower percentage stroma than Luminal A tumors which had the highest stromal proportion (Basal-like stroma: 29.0%, LumA stroma: 44.0%). Luminal B and HER2-enriched also had significantly lower percentage stroma [−8.8% (95% CI: −13.8, −3.8) and −14.9% (95% CI: −17.8, −12.0), respectively] than Luminal A tumors (Fig. 3). Lower stromal area also was observed among tumors with high PAM50-derived risk of recurrence scores (ROR-PT) and high combined histologic grade. When tumors with low versus high ROR-PT were compared, the mean difference in stromal area was −20.5% (95% CI: −24.3, −16.7). Tumors with high combined histologic grade had substantially less stroma than either low- or intermediate-grade tumors [28.4%, mean difference (high vs low grade) = −19.3% (95% CI: −21.9, −16.7)]. Finally, invasive lobular carcinomas were associated with substantially more stroma (53.2%) than tumors with ductal histology [38.1%, mean difference = −15.1% (95% CI: −18.3, −12.0)].

3.2.2. Immune infiltrate

The presence of immune infiltrate (≥1% of TMA core area) was associated with aggressive disease phenotypes (Table 3, Fig. 2B). Among IHC subtypes, immune infiltrate was least common among HR-positive breast cancers (33.6%), and more common among triple-negative [53.2%, RFD = 19.6% (95% CI: 13.3, 25.8)] and HER2-positive tumors [41.6%, RFD = 8.0% (95% CI: 0.1, 15.2)]. Similarly, among PAM50 subtypes, immune infiltrate was least common in Luminal A tumors (30.2%) and was significantly associated with Basal-like tumors [55.2%, RFD = 25.0% (95% CI: 17.9, 32.0)]. Associations were also significant between immune infiltrate and Luminal B or HER2-enriched tumors. Finally, the presence of an immune infiltrate was associated with both high (49.2%) and intermediate ROR-PT scores (38.1%) when compared to tumors with low ROR-PT scores [23.8%, High RFD = 25.4% (95% CI: 16.2, 34.4); Intermediate RFD = 14.3% (95% CI: 7.1, 21.2)]. Similarly, both intermediate and high combined histologic grade were significantly associated with immune infiltrate. Intermediate [35.8%, RFD = 10.5% (95% CI: 4.2, 16.6)] and high-grade tumors [50.4%, RFD = 25.1% (95% CI: 18.8, 31.3)] had high amounts of immune infiltrate relative to low grade tumors (25.3%). Among histological subtypes, lobular tumors (27.5%) were significantly less likely to contain immune infiltrate than ductal tumors [38.4%, RFD = 10.9% (95% CI: 3.4, 18.0)].

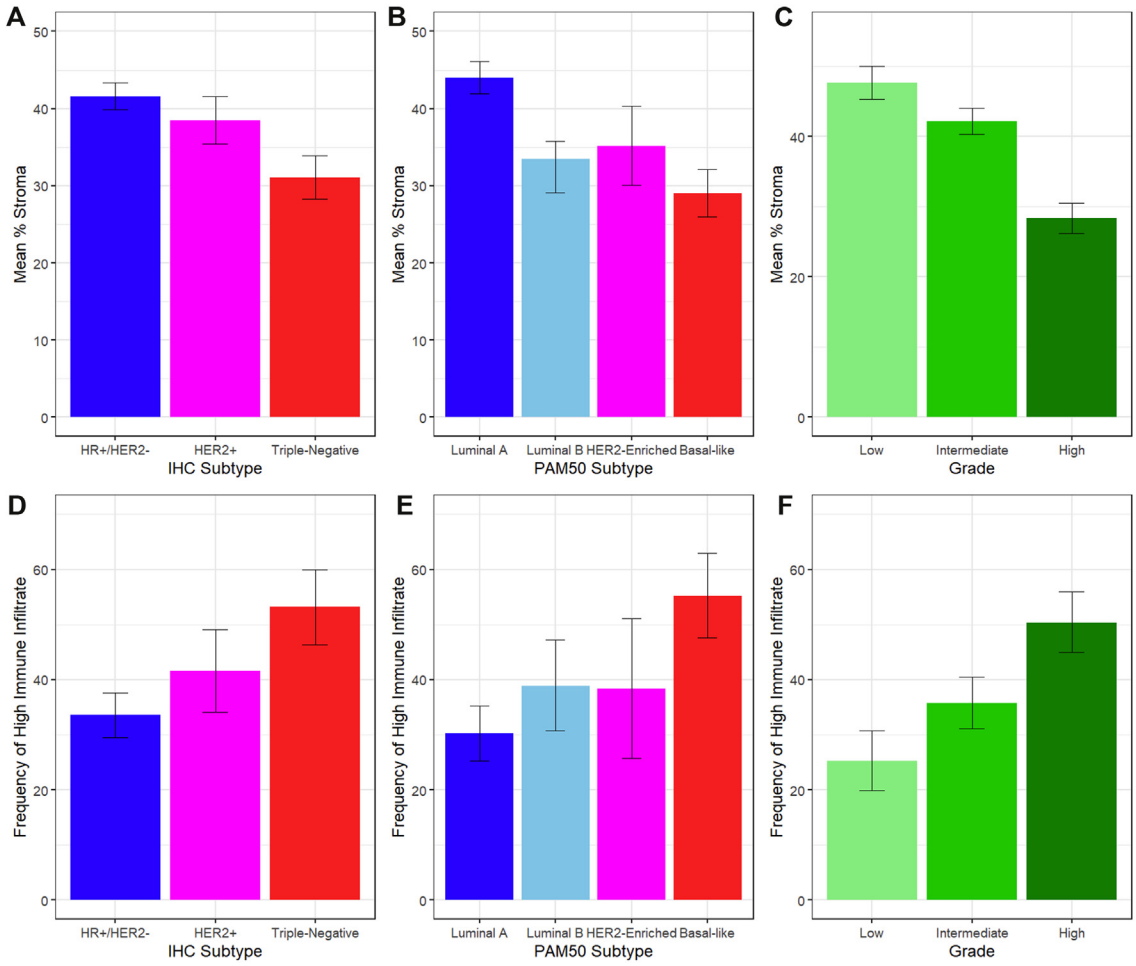

Fig. 2 Lower percentages of stroma are associated with more aggressive disease phenotypes such as triple-negative (A) or Basal-like (B) breast cancer subtype and higher grade (C). The presence of immune infiltrate is associated with more aggressive disease phenotypes such as triple-negative (D) or Basal-like (E) breast cancer subtype and higher grade (F).

3.2.3. Stromal composition: Basal-like versus non-Basal-like tumors

An exploratory ROC analysis was performed to determine the potential for stromal proportion in TMA cores to distinguish breast cancer subtypes (Fig. 3). Among all tumors, the stromal proportion yielded an AUC of 0.680 for distinguishing Basal-like from non-Basal-like tumors, with a cut-point of <28.4% identifying more Basal-like tumors. Among all tumors that were classified as HR+/HER2-by IHC, the stromal proportion yielded an AUC of 0.662 with an optimal cut-point of 22.0% stroma for distinguishing Luminal A from non-Luminal A tumors based on PAM50 results.

4. Discussion

Using an efficient, high-throughput computational algorithm, we found that clinical and molecular characteristics of breast cancer are associated with distinct tumor microenvironments. In particular, we observed decreasing proportions of stromal content and the presence of at least 1% lymphocytic immune infiltrate in more aggressive tumors (Basal-like, triple-negative, high ROR-PT scores, high-grade). These results were consistent across two digital image analysis platforms, Aperio Genie and Definiens. The proportion of stroma was a strong predictor of subtypes of breast cancer in exploratory ROC analysis. The study population, CBCS, oversamples young and Black women and therefore we had good sample size to detect features of aggressive and Basal-like tumors that are more common in these groups. Given the preponderance of hormone receptor-positive breast cancer in the general population, the potential to separate Luminal A versus Luminal B tumor may have the highest potential clinical impact. Automated methods could be used to prescreen ER + tumors for further evaluation for mitotic counts or genomic testing.

Increasingly, computational pathology methods are being applied to histopathological images to elucidate

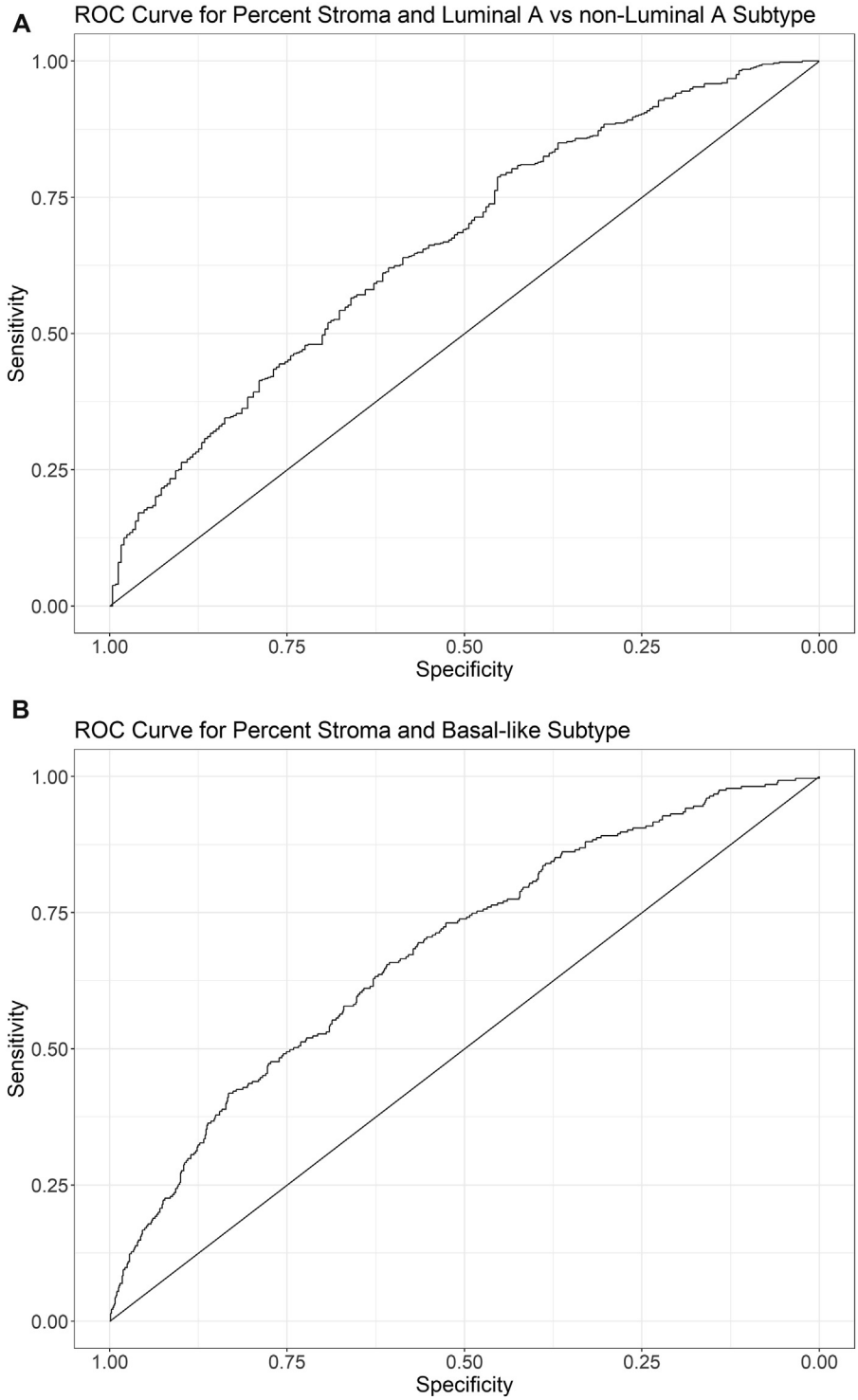

Fig. 3 The percentage of stroma present in a tumor is able to roughly distinguish between Luminal A and Luminal B tumors (A), as well as between Basal-like and non-Basal-like tumors (B).

disease heterogeneity and outcomes [8,9]. High-throughput computational techniques to assess morphological features may improve upon manual assessment by being more efficient, objective, and consistent than a human observer [11,12]. A study by Beck at al [9]. observed that stromal features extracted from breast cancer TMAs using machine learning techniques were strongly associated with overall survival among breast cancer patients. In that study, models including three of the most predictive stromal features were stronger predictors of overall survival than epithelial

Table 3 Associations between lymphocytic immune infiltrate (dichotomized as <1% vs ≥ 1%) and tumor and clinical characteristics in the study population.

	Low Immune (<1%) (n = 875) N (%)	High Immune (≥1%) (n = 705) N (%)	RFD (95% CI)[a]	*p*-value[b]
IHC Subtype				
HR + HER2-	688 (73.5)	426 (59.3)	REF	
HER2+	118 (12.6)	105 (14.6)	8.0 (0.9, 15.2)	0.0843
Triple-Negative	130 (13.9)	187 (26.0)	19.6 (13.3, 25.8)	<0.0001
Missing	19	16		
PAM50 Subtype[c]				
Luminal A	403 (60.8)	205 (42.6)	REF	
Luminal B	111 (16.7)	84 (17.5)	8.7 (0.9, 16.7)	0.0843
HER2-enriched	38 (5.7)	28 (5.8)	8.2 (−3.8, 20.7)	0.2001
Basal-like	111 (16.7)	164 (34.1)	25.0 (17.9, 32.2)	<0.0001
Normal-like	14	8	NA	NA
Missing	278	245		
ROR-PT Score				
Low	157 (23.5)	60 (12.4)	REF	
Intermediate	404 (60.6)	296 (61.2)	14.3 (7.1, 21.2)	0.0003
High	106 (15.9)	128 (26.4)	25.4 (16.2, 34.4)	<0.0001
Missing	288	250		
Grade				
Low	239 (25.2)	93 (12.9)	REF	
Intermediate	384 (40.5)	233 (33.6)	10.5 (4.2, 16.6)	0.0040
High	324 (34.2)	372 (53.5)	25.1 (18.8, 31.3)	<0.0001
Missing	8	7		
Morphology				
Ductal	826 (87.2)	670 (91.9)	REF	
Lobular	121 (12.8)	59 (8.1)	−10.9 (−18.0, −3.4)	0.0134
Missing	8	5		

[a] Models adjusted for age and race ($p < 0.05$)
[b] *P*-values adjusted for multiple testing using Bonferroni-Holm method.
[c] Normal-like samples were excluded for the purposes of PAM50 assessment because of small sample size.

features alone [9]. The authors observed that tumors with larger contiguous areas of stroma were associated with better prognosis, consistent with our observations that a higher stromal proportion is associated with HR-positive, low-grade tumors.

Immune cells are known to be an important component of the tumor microenvironment [13−15]. Our findings that immune infiltrates are more common in triple-negative and HER2-expressing breast cancer subtypes is consistent with previous findings [16,17]. Previous studies using data from the Cancer Genome Atlas (TCGA) demonstrated high immune gene expression among Basal-like and HER2-enriched breast cancers [13]. Furthermore, tumors with higher proportions of TILs have been shown to be associated with response to neoadjuvant therapy and survival among women with triple-negative and HER2-positive breast cancer [16,18].

Strengths of this study include the diverse collection of histological specimens in CBCS that allows for comparisons across a wide range of disease phenotypes [3]. In particular, CBCS recruited a large number of Black women, an often-underrepresented group in molecular and pathologic studies. The study also included molecular profiling by RNA-based assays, which allowed assessment of RNA-based and protein-based tumor features in relation to tissue composition, as well as clinical pathologic features [4]. A potential limitation of digital approaches is misclassification of some image features. However, we do not anticipate that classification error would vary by tumor subtype. Similar to Beck et al. [9], we analyzed breast cancer TMAs, which may result in some differences in composition estimates relative to what is observed in whole slides. Cores tend to be sampled centrally in a tumor, and some studies suggest that immune infiltration is most common on the tumor periphery [19]. If this is true, we may have underestimated the association between peripheral immune infiltrate and tumor aggressiveness, however our estimates are likely reliable for assessing intratumoral infiltrates.

Recent advances in breast cancer immune-oncology and computational approaches to histologic image analysis highlight the importance of immune cells and stromal features in the tumor microenvironment [9,20,21]. Our results confirmed significant differences in both immune cell

infiltration and stromal content among breast cancer subtypes, with increased immune cells and lower stromal content in tumors with aggressive phenotypes. The integration of digital image analysis of TMAs with clinical and molecular data provides a powerful, high-throughput approach to the study of breast cancer in large, population-based cohorts.

Acknowledgements

The Carolina Breast Cancer Study was supported by a grant from UNC Lineberger Comprehensive Cancer Center, which is funded by the University Cancer Research Fund of North Carolina, the Susan B Komen Foundation (OGUNC1202), the National Cancer Institute of the National Institutes of Health (P01CA151135), and the National Cancer Institute Specialized Program of Research Excellence (SPORE) in Breast Cancer (NIH/NCI P50-CA58223). This research recruited participants &/or obtained data with the assistance of Rapid Case Ascertainment, a collaboration between the North Carolina Central Cancer Registry and UNC Lineberger. RCA is supported by a grant from the National Cancer Institute of the National Institutes of Health (P30CA016086). The authors would like to acknowledge the University of North Carolina BioSpecimen Processing Facility for sample processing, storage, and sample disbursements (http://bsp.web.unc.edu/). We are grateful to CBCS participants and study staff.

References

[1] 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. Hum Pathol 2016;48:60−8.

[2] Chollet-Hinton L, Puvanesarajah S, Sandhu R, Kirk EL, Midkiff BR, Ghosh K, et al. Stroma modifies relationships between risk factor exposure and age-related epithelial involution in benign breast. Mod Pathol 2018.

[3] Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995;35:51−60.

[4] Troester MA, Sun X, Allott EH, Geradts J, Cohen SM, Tse CK, et al. Racial differences in PAM50 subtypes in the Carolina breast cancer study. J Natl Cancer Inst 2018;110.

[5] Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. Mod Pathol 2018;31:1502−12.

[6] Mittal S, Stoean C, Kajdacsy-Balla A, Bhargava R. Digital assessment of stained breast tissue images for comprehensive tumor and microenvironment analysis. Front Bioeng Biotechnol 2019;7:246.

[7] Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing 2016;191:214−23.

[8] Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer 2018;4:30.

[9] Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011;3:108ra13.

[10] Rabe K, Snir OL, Bossuyt V, Harigopal M, Celli R, Reisenbichler ES. Interobserver variability in breast carcinoma grading results in prognostic stage differences. Hum Pathol 2019;94:51−7.

[11] Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. Annu Rev Pathol 2013;8:331−59.

[12] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019;16:703−15.

[13] Iglesia MD, Parker JS, Hoadley KA, Serody JS, Perou CM, Vincent BG. Genomic analysis of immune cell infiltrates across 11 tumor types. J Natl Cancer Inst 2016;108.

[14] Mittal S, Brown NJ, Holen I. The breast tumor microenvironment: role in cancer development, progression and response to therapy. Expert Rev Mol Diagn 2018;18:227−43.

[15] Place AE, Jin Huh S, Polyak K. The microenvironment in breast cancer progression: biology and implications for treatment. Breast Cancer Res 2011;13:227.

[16] Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol 2018;19:40−50.

[17] Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol 2015;26:259−71.

[18] Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. J Clin Oncol 2015;33:983−91.

[19] Konig L, Mairinger FD, Hoffmann O, Bittner AK, Schmid KW, Kimmig R, et al. Dissimilar patterns of tumor-infiltrating immune cells at the invasive tumor front and tumor center are associated with response to neoadjuvant chemotherapy in primary breast cancer. BMC Cancer 2019;19:120.

[20] Emens LA, Adams S, Cimino-Mathews A, Disis ML, Gatti-Mays ME, Ho AY, et al. Society for Immunotherapy of Cancer (SITC) clinical practice guideline on immunotherapy for the treatment of breast cancer. J Immunother Cancer 2021;9.

[21] Walens A, Olsson LT, Gao X, Hamilton AM, Kirk EL, Cohen SM, et al. Protein-based immune profiles of basal-like vs. luminal breast cancers. Lab Invest 2021 Jun;101(6):785−93. https://doi.org/10.1038/s41374-020-00506-0. Epub 2021 Feb 23. PMID: 33623115; PMCID: PMC8140991.

[22] Fuller AM, Yang L, Hamilton AM, Pirone JR, Oldenburg AL, Troester MA. Epithelial p53 status modifies stromal-epithelial interactions during basal-like breast carcinogenesis. J Mammary Gland Biol Neoplasia 2021;26:89−99. https://doi.org/10.1007/s10911-020-09477-w.

[23] Camp JT, Elloumi F, Roman-Perez E, Rein J, Stewart DA, Harrell JC, et al. Interactions with fibroblasts are distinct in Basal-like and luminal breast cancers. Mol Cancer Res 2011;9:3−13. https://doi.org/10.1158/1541-7786.MCR-10-0372.

[24] Casbas-Hernandez P, D'Arcy M, Roman-Perez E, Brauer HA, McNaughton K, Miller SM, et al. Role of HGF in epithelial-stromal cell interactions during progression from benign breast disease to ductal carcinoma in situ. Breast Cancer Res 2013;15:R82. https://doi.org/10.1186/bcr3476.