Contribution of Prediagnostic Host Factors to Shaping the Stromal Microenvironment of Breast Cancer among Sub-Saharan African Women

Mustapha Abubakar[1], Thomas U. Ahearn[1], Maire A. Duggan[2], Scott Lawrence[3], Ernest K. Adjei[4], Joe-Nat Clegg-Lamptey[5], Joel Yarney[5], Beatrice Wiafe-Addai[6], Baffour Awuah[4], Seth Wiafe[7], Kofi Nyarko[8], Francis S. Aitpillah[4], Daniel Ansong[9], Stephen M. Hewitt[10], Louise A. Brinton[1], Jonine D. Figueroa[1,11], Montserrat Garcia-Closas[1], Lawrence Edusei[5], and Nicolas Titiloye[4]; for the Ghana Breast Health Study Team

ABSTRACT

Background: The stromal microenvironment (SME) is integral to breast cancer biology, impacting metastatic proclivity and treatment response. Emerging data indicate that host factors may impact the SME, but the relationship between prediagnostic host factors and SME phenotype remains poorly characterized, particularly among women of African ancestry.

Methods: We conducted a case-only analysis involving 792 patients with breast cancer (17–84 years) from the Ghana Breast Health Study. High-accuracy machine-learning algorithms were applied to standard H&E-stained images to characterize SME phenotypes [including percent tumor-associated connective tissue stroma, Ta-CTS (%); tumor-associated stromal cellular density, Ta-SCD (%)]. Associations between prediagnostic host factors and SME phenotypes were assessed in multivariable linear regression models.

Results: Decreasing Ta-CTS and increasing Ta-SCD were associated with aggressive, mostly high-grade tumors (P-value < 0.001). Several prediagnostic host factors were associated with Ta-SCD independently of tumor characteristics. Compared with nulliparous women, parous women had higher levels of Ta-SCD [mean (standard deviation, SD) = 31.3% (7.6%) vs. 28.9% (7.1%); P-value = 0.01]. Similarly, women with a positive family history of breast cancer had higher levels of Ta-SCD than those without family history [mean (SD) = 33.0% (7.5%)] vs. 30.9% (7.6%); P-value = 0.03]. Conversely, increasing body size was associated with decreasing Ta-SCD [mean (SD) = 31.6% (7.4%), 31.4% (7.3%), and 30.1% (8.0%) for slight, average, and large body sizes, respectively; P-value = 0.005].

Conclusions: Epidemiological risk factors were associated with varying degrees of stromal cellularity in tumors, independently of clinicopathological characteristics.

Impact: The findings raise the possibility that epidemiological risk factors may partly influence tumor biology via the stromal microenvironment.

Introduction

In addition to the neoplastic parenchymal cells, tumors are composed of a complex mixture of cellular and noncellular elements that collectively comprise the stromal microenvironment (1–4). The stromal microenvironment plays a critical role in tumor initiation, progression, and response to treatment (1, 5–8). Although widely believed to evolve in parallel with the neoplastic parenchyma, accumulating data suggest that changes in the stromal microenvironment may predate breast cancer development and that premalignant/preinvasive stromal changes could influence the biologic phenotype of ensuing tumors (9–11).

Emerging evidence suggest that individual factors, such as parity, body mass index (BMI), and race and ethnicity may impact the stromal microenvironment (12–15). This is of particular relevance given the observed heterogeneity in the incidence of breast cancer subtypes according to individual factors (16, 17). For instance, parity has been shown in several epidemiological studies to be more strongly associated with risk of triple negative (TN) breast cancer (TNBC), an aggressive form of breast cancer that is characterized by the lack of expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). On the other hand, postmenopausal obesity has been shown to be more strongly associated with elevated risk of high grade/ER$^+$ tumors (18–21). Other risk factors, such as early menarche, nulliparity, lack of breastfeeding, postmenopausal status, and

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland. [2]Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Canada. [3]Molecular and Digital Pathology Laboratory, Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick, Maryland. [4]Komfo Anokye Teaching Hospital, Kumasi, Ghana. [5]Korle Bu Teaching Hospital, Accra, Ghana. [6]Peace and Love Hospital, Kumasi, Ghana. [7]Loma Linda University, School of Public Health, Loma Linda, California. [8]University of Ghana, Accra, Ghana. [9]Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [10]Center for cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland. [11]Usher Institute and Cancer Research UK Edinburgh Centre, University of Edinburgh, Edinburgh, United Kingdom.

M. Garcia-Closas, L. Edusei, and N. Titiloye contributed equally to this article.

Ghana Breast Health Study Team members listed in the acknowledgments

Corresponding Author: Mustapha Abubakar, Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20850. E-mail: mustapha.abubakar2@nih.gov

Cancer Epidemiol Biomarkers Prev 2024;XX:XX–XX

doi: 10.1158/1055-9965.EPI-24-0390

use of menopausal hormone therapy have been shown to be more consistently associated with elevated risk of ER^+ than ER^- breast cancer subtypes (16, 18, 22). Furthermore, compared with women of European ancestry, women of African ancestry tend to have a higher prevalence of aggressive, TN or high grade, breast cancer phenotypes (23–25). Similar patterns of higher rates of aggressive breast cancer phenotypes are seen among women in sub-Saharan Africa (26, 27).

Although most previous studies of breast cancer etiologic heterogeneity have focused on tumor parenchymal characteristics such as ER, PR, and HER2, distinct stromal microenvironment phenotypes may contribute to observed differences in tumor biology by host factors (12, 13, 15, 28). In general, stromal changes reminiscent of chronic inflammation and wound repair have been shown to predominate among parous than nulliparous, obese than normal weight, and African American than White women (15, 28). In particular, and with respect to the latter, the densities of tumor-associated macrophages, endothelial cells, and microvessels in the stromal microenvironment, in addition to the expression of an interferon signature, have been shown to be higher among women of African than European ancestry after accounting for age, tumor stage, ER subtype, and grade (15). To date, however, there have been no studies investigating the impact of host factors on the stromal microenvironment of breast cancer among women from an indigenous sub-Saharan African population who are more likely to develop aggressive breast cancer subtypes and less likely to undergo screening.

Our main aim in this study was to investigate the association between host factors and stromal microenvironment phenotypes among women with breast cancer from a sub-Saharan African population. To address this aim, we utilized high-accuracy machine learning algorithms to characterize the stromal microenvironment of primary invasive breast cancer using digitized hematoxylin and eosin (H&E)-stained sections of tumor tissues. Two biologically and prognostically relevant H&E-based stromal microenvironment phenotypes were considered, including the amount of tumor-associated connective tissue stroma (i.e., Ta-CTS, which is defined as the percentage of the total tumor area that is connective tissue stroma relative to the tumor neoplastic parenchyma; refs. 29, 30); and the tumor-associated stromal cellular density [i.e., Ta-SCD, which is a metric of the proportion of the tumor-stroma that is occupied by nucleated cells such as immune cells [e.g., tumor infiltrating lymphocytes (TIL)], macrophages, and natural killer (NK) cells], and other cellular components such as fibroblasts, endothelial cells, and pericytes (31). Relationships between breast cancer risk factors and the stromal microenvironment phenotypes (i.e., Ta-CTS and Ta-SCD) were assessed overall and according to relevant clinicopathologic characteristics, including age, ER status, and histologic grade.

Materials and Methods

Study population

The Ghana Breast Health Study (GBHS) is a multidisciplinary, population-based, case-control study in Ghana, details of which have been previously described (32). In brief, cases were women who presented with lesions suspicious for breast cancer in three hospitals in Kumasi (Komfo Anokye Teaching Hospital and Peace and Love Hospital) and Accra (Korle Bu Teaching Hospital), Ghana, between 2013 and 2015. These hospitals represent the primary hospitals that provide surgical and oncological care for patients with breast cancer in Ghana (32). A total of 1,126 patients with breast cancer were recruited as part of the GBHS (33). Of these, 792 had available digitized H&E-stained images that were confirmed by study pathologists to contain invasive breast cancer. Accordingly, the current analysis comprised 792 patients aged 17 to 84 years with histologically confirmed invasive breast cancer and for whom we had digitized H&E-stained images. The study was approved by institutional review boards at the National Institutes of Health (NIH; Institutional Review Board of the National Cancer Institute), Kwame Nkrumah University of Science and Technology (Kumasi, Ghana), Noguchi Memorial Institute for Medical Research (Accra, Ghana), School of Medical Sciences, Komfo Anokye Teaching Hospital (Kumasi, Ghana), and Westat (Rockville, MD, USA). This study was conducted in accordance with the Declaration of Helsinki ethical guidelines. All study participants provided written informed consent.

Risk factor information

Data on established breast cancer risk factors, including demographic factors, menstrual and reproductive characteristics, family history of breast cancer, medical history, occupational history, anthropometric, and physical activity variables, were obtained using a detailed questionnaire that was administered to each participant in the hospital by trained personnel. The women were asked specifically about pregnancy and breastfeeding practices, including number of childbirths and total duration of breastfeeding for each birth. Data on average lifetime body size were obtained using pictograms that were classified as slight, average, slightly heavy, and heavy (32). This study relied on pictograms due to lack of scales in some places and to overcome the challenge of weight loss due to disease. For this analysis, we used risk factor variables that have been previously described and shown to be associated with breast cancer risk in this population (32, 33). Participants' ages (years) were categorized as <35, 35 to 44, 45 to 54, and ≥55 years. Age at menarche (years) was categorized as <15, 15, 16, ≥17. Parity was classified as nulliparous or parous (≥1 child). Among parous women, number of children was categorized as 1, 2, 3, 4, and ≥5 children. Age at first birth (years) was categorized as <19, 19 to 21, 22 to 25, and ≥26. Breastfeeding (median, per birth, months) was categorized as <13, 13 to 18, and ≥19. Body size phenotypes were classified as slight, average, or heavy (including slightly heavy). Family history was considered as positive if a patient had a history of breast cancer in a first degree relative and negative if there was no such history.

Clinicopathological data

Pretreatment core-needle biopsies (n = 4–8) from each patient were fixed in 10% neutral buffered formalin for 24 to 72 hours and subsequently processed into paraffin-embedded tissue blocks (32). Data on tumor size were based on clinical palpation of the mass at the time of diagnosis (34). Information on histologic grade was obtained through centralized pathologist (MAD) review (35). Immunohistochemical (IHC) staining for ER, PR, HER2 was performed using standard laboratory protocols (33). Tumors were considered to be ER^+ and/or PR^+ if ≥10% of the tumor cells demonstrated positive staining. A staining of 3+ for HER2 was considered as positive while borderline and negative cases were considered as HER2 negative. As previously reported (33), good agreements (79%, 65%, and 78% for ER, PR, and HER2, respectively; $P < 0.01$) were observed between scores that were determined by pathologists in Ghana and at the NCI. Breast cancer subtypes were defined using data on ER, PR, and HER2 as follows: luminal A-like

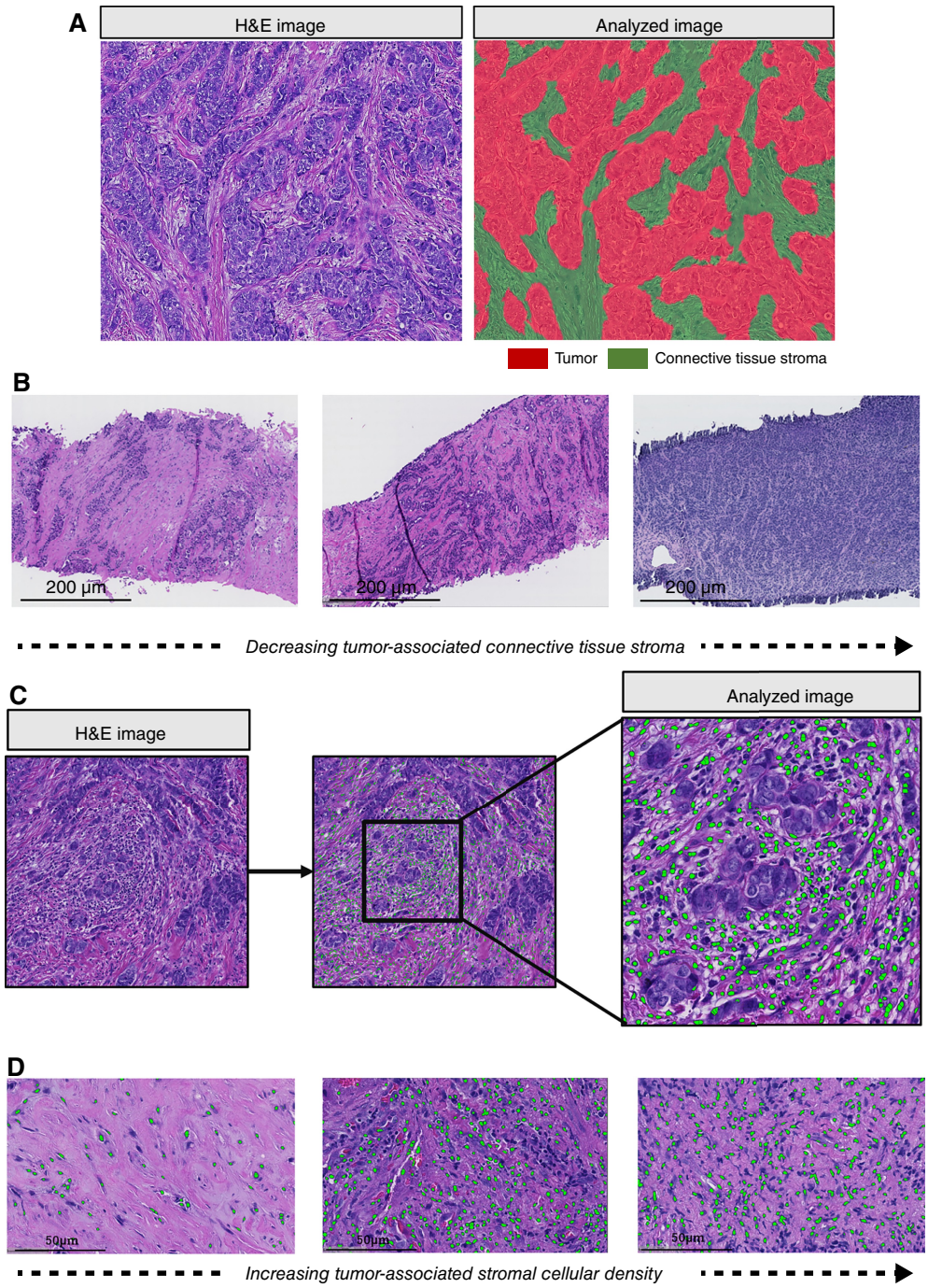

Figure 1.

Machine-learning classification of Ta-CTS and Ta-SCD. Digitized H&E-stained images (**A**) were analyzed using optimized machine-learning scripts based on the random forest algorithm. A 160-datapoint random forest tissue-classifier algorithm was trained and optimized to identify, segment, and quantify (in mm^2) areas on H&E images consisting of tumor (94 datapoints, red regions on analyzed image) and stroma (66 datapoints; green regions on analyzed image). **B,** shows examples of H&E images from patients with varying amounts of Ta-CTS based on machine-learning image analysis [high to low Ta-CTS is shown from left to right, (**B**)]. Stromal regions on digitized H&E-stained images (**C**) were digitally annotated to allow subsequent cell detection to be confined to the stromal compartment. The cell detection script was parameterized to detect nonmalignant nucleated cells in the stroma (green dots in analyzed image) based on size, shape, nuclear detection weight, nuclear contrast threshold, nuclear optical density, etc. (Supplementary Table S1). The algorithm was trained to exclude infiltrating nests of malignant epithelial cells from the Ta-SCD metric (analyzed image). The cell detection script showed strong correlation with two pathologists' (P1 and P2) manual cell counts within the intratumoral (R = 0.93 and 0.75 vs. P1 and P2, respectively) and peritumoral (R = 0.74 and 0.73 vs. P1 and P2, respectively) stromal compartments. **D,** shows examples of patches of H&E images from patients with varying amounts of Ta-SCD-based on machine-learning image analysis [low to high Ta-SCD is shown from left to right, (**D**)].

($ER^+/PR^+/HER2^-$); luminal B-like (ER^+ and/or PR^+ and $HER2^+$ or $ER^+/PR^-/HER2^-$); HER2-enriched ($ER^-/PR^-/HER2^+$); and TNBC ($ER^-/PR^-/HER2^-$).

Machine learning characterization of tumor-associated stromal microenvironment phenotypes on H&E images

H&E-stained sections were digitized at the NCI and archived using the Halo Link digital image repository (Indica Labs, Albuquerque, NM). Imaging quality control was performed by pathologists (NT, LE, MA) using a standard operating protocol designed to identify and positively annotate tumor regions on the slide (including intratumoral and peritumoral stroma) and to negatively annotate regions with substantial crushing artifacts, damaged tissue, or widespread necrosis. In addition, the pathologists used a computer-assisted visual counting tool to identify and count nucleated cells within well-defined regions ($500 \times 500\ \mu m^2$) of stroma.

Digitized images were analyzed using optimized machine-learning scripts based on the random forest algorithm. First, a 160-datapoint random forest tissue-classifier script was trained and optimized to identify, segment, and quantify (in mm^2) areas on each image consisting of tumor (94 datapoints) and stroma (66 datapoints; **Fig. 1A** and **B**). The performance of the tissue classifier script in distinguishing between tumor and stroma was confirmed by visual inspection of random images by study pathologists. As previously reported in an independent study population (36), the random forest tissue-classifier demonstrated excellent reproducibility when comparing scripts that were independently trained by two pathologists (Spearman's rho = 0.95 and 0.97 for epithelium and stroma, respectively; ref. 36). Stromal regions were digitally annotated to allow subsequent cell detection to be confined to the stromal compartment. Next, a previously validated (31) cell-detection script was reparametrized (Supplementary Table S1) to identify and count nucleated cells in the stroma including lymphocytes, macrophages, fibroblasts, endothelial cells, etc. (**Fig. 1C** and **D**). In validation analyses, the cell detection script showed strong correlation with two pathologists' manual cell counts within the intratumoral ($R = 0.93$ and 0.75 *vs.* P1 and P2, respectively) and peritumoral ($R = 0.74$ and 0.73 *vs.* P1 and P2, respectively) stromal compartments. Optimized scripts were used for centralized image analysis, which was blinded to patients' demographic, epidemiological, or clinical characteristics.

Percent Ta-CTS was calculated by dividing the connective tissue area (mm^2) on the slide by the fibroglandular tissue area (i.e., which combines regions of connective tissue and tumor parenchyma) on the slide and multiplying by 100. Ta-SCD was calculated by dividing the total number of cells in the stroma by the total stromal area (mm^2). This was converted to a percentage by multiplying the total number of nucleated cells by the average area (mm^2) of a single nucleus (2.0×10^{-04}), dividing this by the total stromal area (mm^2), and multiplying by 100. There was near-perfect positive correlation ($r = 0.99$; Supplementary Fig. S1) between standard and percent Ta-SCD. The latter was used for all further analysis to facilitate interpretability.

Statistical analysis

The Anova and Kruskal–Wallis tests were used to evaluate differences in the distributions of Ta-CTS and Ta-SCD by patients' characteristics. Multivariable linear regression models were used to test associations between risk factors, patients' clinicopathological characteristics, and stromal microenvironment phenotypes. Ta-CTS and Ta-SCD were normally distributed hence generalized linear regression was used to test associations with risk factors and tumor characteristics. In separate linear regression models, Ta-CTS and Ta-SCD were modeled as outcomes while the risk factors and patients' characteristics were modeled as predictors. Partially adjusted models contained each relevant risk factor or clinicopathological characteristic in addition to age, study site, and tissue area. The multivariable models were mutually adjusted for all the risk factors or clinicopathological factors that were considered as well as age, study site, and tissue area. Multiplicative interaction terms were included in full models to test for evidence of effect modification between host factors and tumor characteristics in relation to Ta-CTS and Ta-SCD. Missing covariate values on tumor characteristics were imputed using the multiple (×5) imputation by chained equations (MICE) approach (37) with appropriate variance adjustment by Rubin's Formula (38) for all analyses. The frequency of missingness for tumor characteristics ranged from ~5% for tumor size to 29% for HER2. It was decided, a priori, to not impute tumor characteristics with more than 30% missingness. Data on risk factors were more complete, with frequency of missingness ranging from 0.2% for parity to 13% for age at menarche. The other risk factors had <6% missingness. All analyses were two-sided and were performed using Stata statistical software version 16.1. P values < 0.05 were considered statistically significant.

Data availability

All the datasets used and/or analyzed during the current study are available upon request from the corresponding author.

Results

Description of study population

Of the 792 participants, two had missing information on age and were dropped from all further analyses. The final analytic population comprised 790 patients aged 17 to 84 (mean = 49.8, median = 49.0) years at diagnosis. Fifty-two percent of the patients had ER^+ and/or PR^+ tumors, while 24% had $HER2^+$ disease. The majority (~70%) of the patients had high (grade 3) disease and about 6% had low (grade 1) grade tumors. Ninety-seven percent of the tumors in this population were at least 2 cm in size at diagnosis. The distributions of breast cancer subtypes were 30%, 34%, 8%, and 28% for luminal A-like, luminal B-like, HER2-enriched, and TNBC, respectively. The median age at menarche was 15 (range = 9–22) years. About 91% of the population had at least one child at the time of breast cancer diagnosis, with the majority having three or more children. Thirty-two percent of the population breastfed for <13 months/pregnancy while 13% breastfed for at least 19 months/pregnancy. The distributions of body size phenotypes were 25%, 40%, and 35% for slight, average, and heavy. Fifty-five patients (~7%) reported having a positive family history (**Table 1**).

Associations of clinicopathological factors with stromal microenvironment phenotype

The mean, median (range) of Ta-CTS (%) and Ta-SCD (%) were 76.2, 78.9 (13.9–99.6), and 31.0, 30.1 (10.7–58.7), respectively (**Table 2**). In crude analysis, the distributions of Ta-CTS and Ta-SCD differed significantly by histologic grade (**Table 2**), with Ta-CTS levels decreasing with increasing grade ($P < 0.0001$) and Ta-SCD levels increasing with increasing grade levels ($P < 0.0001$). We also observed the distributions of Ta-CTS to vary by ER and

Table 1. Description of clinicopathologic and risk factor characteristics of 792 patients with breast cancer participating in the GBHS.

Characteristic	Frequency (%)
Overall	792
Age, years	
<35	84 (10.6)
35–44	198 (25.1)
45–54	231 (29.2)
≥55	277 (35.1)
Missing	2
ER status	
Negative	287 (48.3)
Positive	307 (51.7)
Missing	198
PR status	
Negative	282 (47.7)
Positive	309 (52.3)
Missing	201
HER2 status	
Negative	422 (75.5)
Positive	137 (24.5)
Missing	233
Grade	
Low	38 (5.6)
Intermediate	168 (24.7)
High	474 (69.7)
Missing	112
Size	
<2 cm	23 (3.0)
2–5 cm	248 (32.9)
>5 cm	483 (64.1)
Missing	38
Subtype	
Luminal A-like	166 (30.0)
Luminal B-like	188 (33.9)
HER2-enriched	47 (8.4)
TNBC	154 (27.7)
Missing	237
Age at Menarche, years	
<15	185 (27.0)
15	183 (26.7)
16	151 (22.0)
≥17	167 (24.3)
Missing	106
Parity	
Nulliparous	71 (9.0)
Parous	719 (91.0)
Missing	2
Number of children	
1	88 (12.2)
2	125 (17.4)
3	136 (18.9)
4	123 (17.1)
≥5	247 (34.4)
Missing	2
Age at first birth (years)	
<19	230 (30.8)
19–21	200 (26.8)
22–25	180 (24.1)
≥26	137 (18.3)
Missing	45
Breastfeeding, months	
<13	240 (31.9)
13–18	415 (55.2)
≥19	97 (12.9)
Missing	40
Joint parity/breastfeeding (months)	
Nulliparous	71 (9.4)
Parous/<13	169 (22.5)
Parous/13–18	415 (55.2)
Parous/≥19	97 (12.9)
Missing	40
Body Size	
Slight	190 (25.5)
Average	296 (39.7)
Heavy	260 (34.8)
Missing	46
Family history	
None	726 (93.0)
Yes	55 (7.0)
Missing	11

(Continued on the following column)

Table 1. Description of clinicopathologic and risk factor characteristics of 792 patients with breast cancer participating in the GBHS. (Cont'd)

HER2 status, with ER^+ and $HER2^+$ tumors having higher Ta-CTS than ER^- and $HER2^-$ tumors, respectively (**Table 2**). In linear regression models partially adjusted for each tumor characteristic as well as age, study site, and tissue area, we found ER, HER2, TNBC, and histologic grade to be associated with Ta-CTS. Following mutual adjustments for all tumor characteristics, ER [β (95% confidence interval, CI)$_{ER^+ \text{ vs. } ER-}$ = 2.91 (0.19, 5.63); P = 0.03], HER2 [β (95% CI)$_{HER2+ \text{ vs. } HER2-}$ = 4.39 (1.43, 7.35); P = 0.005], and histologic grade [β (95% CI) = −2.07 (−6.70, 2.56), −9.19 (−13.40, −4.98) for grades 2 and 3 *vs.* 1, respectively; P-trend <0.001] remained significantly associated with Ta-CTS (**Table 3**). Ta-SCD was higher with increasing grade in partially adjusted models and after adjustment for other clinicopathological factors in multivariable models [β (95% CI) = 3.53 (1.03, 6.03), 5.22 (2.74, 7.71) for grades 2 and 3 *vs.* 1, respectively; P-trend <0.001]. Increasing Ta-SCD was also suggestively associated with increasing tumor size, but this did not attain statistical significance [β (95% CI) = 0.07 (−3.10, 3.23), 1.20 (−1.95, 4.36) for 2 to 5 and >5 vs. <2 cm, respectively; P-trend = 0.06] (**Table 3**).

Associations of risk factors with stromal microenvironment features

The distributions of Ta-SCD (**Table 2**), but not Ta-CTS, varied significantly according to categories of parity, body size, and family history. Ta-SCD was higher among parous [mean (standard deviation) = 31.3% (7.6%)] than nulliparous [mean (standard deviation) = 28.9% (7.1%)] women (P-value = 0.01) and was higher among women with a positive [mean (standard deviation) = 33.0% (7.5%)] than those with no [mean (standard deviation) = 30.9% (7.6%)] family history (P-value = 0.01). Conversely, increasing body size was associated with decreasing Ta-SCD [mean (standard deviation) = 31.6% (7.4%), 31.4% (7.3%), and 30.1% (8.0%) for slight, average, and large body sizes, respectively; P-value = 0.005]. These findings persisted in mutually adjusted multivariable linear regression models (**Table 4**). Specifically, parity [β (95% CI) vs. nulliparous = 2.92 (1.04,

Table 2. Distributions of tumor microenvironment phenotypes (Ta-CTS and Ta-SCD) by patients' clinicopathological and risk factor characteristics in the GBHS.

		Ta-CTS			Ta-SCD		
Characteristic	**Number**	**Mean (SD)**	**Median (range)**	***P* value[a]**	**Mean (SD)**	**Median (range)**	***P* value[a]**
Overall		76.2 (14.4)	78.9 (13.9–99.6)		31.0 (7.61)	30.1 (10.7–58.7)	
Age, years							
<35	84	77.6 (13.4)	77.7 (42.9–99.6)		31.6 (7.4)	31.4 (12.1–51.9)	
35–44	198	76.1 (16.4)	79.9 (13.9–98.8)		30.7 (7.7)	30.1 (13.0–54.3)	
45–54	231	75.7 (14.0)	77.9 (17.3–99.3)		30.3 (7.5)	30.4 (10.7–53.4)	
≥55	277	76.3 (13.9)	79.3 (21.8–99.1)	0.78	31.7 (7.6)	31.6 (12.4–58.7)	0.14
ER status							
Negative	287	75.1 (15.6)	78.0 (17.3–99.8)		31.1 (7.2)	30.7 (12.1–49.7)	
Positive	307	77.5 (12.7)	80.1 (17.9–98.9)	0.03	30.4 (7.7)	30.0 (10.7–58.7)	0.26
PR status							
Negative	282	76.9 (14.4)	79.9 (17.3–99.7)		30.6 (7.6)	29.9 (10.7–51.9)	
Positive	309	75.9 (13.7)	79.0 (17.9–99.1)	0.43	30.8 (7.4)	30.5 (12.4–58.7)	0.77
HER2 status							
Negative	422	75.6 (14.9)	78.7 (17.3–99.1)		30.6 (7.6)	30.2 (10.7–58.7)	
Positive	137	79.1 (11.2)	80.9 (30.7–99.7)	0.01	30.6 (7.4)	30.7 (14.1–49.7)	0.98
Grade							
Low	38	82.9 (11.0)	84.0 (35.9–98.9)		26.2 (6.9)	26.8 (12.4–38.9)	
Intermediate	168	80.6 (10.0)	82.4 (35.8–99.7)		30.4 (7.7)	30.5 (11.2–47.6)	
High	474	73.1 (15.8)	76.4 (13.9–99.1)	<0.0001	32.0 (7.5)	31.9 (10.7–58.7)	<0.0001
Size							
<2 cm	23	79.0 (14.7)	83.4 (31.5–94.3)		30.4 (8.7)	29.9 (13.1–51.9)	
2–5 cm	248	75.4 (15.4)	78.2 (14.0–99.3)		30.1 (7.5)	30.0 (10.7–58.7)	
>5 cm	483	76.5 (13.9)	79.1 (17.9–99.7)	0.42	31.5 (7.5)	31.2 (11.2–54.3)	0.06
Subtype							
Luminal A-like	166	77.2 (12.9)	80.6 (17.9–98.9)		30.8 (7.6)	30.3 (12.4–58.7)	
Luminal B-like	188	76.3 (13.8)	78.6 (30.8–99.1)		30.1 (7.5)	29.8 (10.7–51.9)	
HER2-enriched	47	80.8 (10.8)	82.4 (51.7–99.7)		32.1 (7.7)	31.9 (17.5–49.7)	
TNBC	154	74.5 (16.1)	77.7 (17.3–99.1)	0.04	30.9 (7.3)	30.4 (12.1–49.7)	0.38
Age at Menarche, years							
<15	185	75.1 (14.6)	76.7 (17.8–99.1)		31.5 (7.4)	31.3 (13.6–51.9)	
15	183	77.4 (14.1)	80.8 (17.3–98.8)		30.7 (7.7)	30.5 (10.7–52.5)	
16	151	75.4 (15.4)	79.2 (13.9–95.1)		31.0 (6.8)	30.6 (14.4–48.8)	
≥17	167	77.0 (13.9)	79.9 (18.3–99.1)	0.33	29.7 (7.8)	29.1 (11.2–54.3)	0.17
Parity							
Nulliparous	71	75.5 (13.8)	78.4 (19.3–96.7)		28.9 (7.1)	28.7 (13.1–46.8)	
Parous	719	76.2 (14.5)	79.0 (13.9–99.7)	0.67	31.3 (7.6)	31.0 (10.7–58.7)	0.01
Breastfeeding, months							
<13	240	75.9 (14.3)	78.6 (17.3–99.3)		30.2 (7.4)	29.9 (13.0–52.5)	
13–18	415	76.8 (13.6)	79.3 (18.3–99.1)		31.4 (7.6)	31.0 (10.7–58.7)	
≥19	97	74.4 (17.6)	78.6 (13.9–98.9)	0.30	31.3 (8.0)	31.3 (12.4–49.7)	0.15
Body Size							
Slight	190	76.0 (12.3)	78.4 (30.7–98.1)		31.6 (7.4)	32.0 (12.1–54.3)	
Average	296	75.6 (15.9)	79.0 (17.3–99.3)		31.4 (7.3)	31.3 (13.1–49.7)	
Heavy	260	77.0 (14.2)	80.6 (13.9–99.1)	0.48	30.1 (8.0)	29.0 (10.7–58.7)	0.005
Family history							
None	726	76.2 (14.4)	78.7 (13.9–99.6)		30.9 (7.6)	30.5 (11.2–58.7)	
Yes	55	76.6 (15.1)	81.3 (17.3–99.1)	0.82	33.0 (7.5)	33.4 (10.7–48.2)	0.03

[a]*P* values were based on ANOVA test.

4.81); *P*-value = 0.002] was associated with higher Ta-SCD after accounting for the other risk factors, but there was no trend with increasing number of births among parous women (**Table 4**). There were no associations between Ta-SCD and age at first birth or breastfeeding. Further, breastfeeding duration did not appear to attenuate the association between parity and Ta-SCD [β (95% CI) $_{\text{parous/breastfeeding vs. nulliparous}}$ = 2.48 (0.37, 4.61); *P*-value = 0.02; 3.34 (1.39, 5.30); *P*-value = 0.001; and 3.36 (0.93, 5.78); *P*-value = 0.007 for parous/breastfeeding <13 months, parous/breastfeeding~ 13–18 months, parous/breastfeeding ≥19 months, respectively]. Having a positive family history [β (95% CI) $_{\text{vs. no FHBC}}$ = 2.36 (0.28, 4.44); *P*-value = 0.02] was also associated with higher Ta-SCD, while increasing body size was inversely associated with Ta-SCD [β (95% CI) = −0.68 (−2.07, 0.72), −2.14 (−3.56, −0.72) for average and heavy *vs.* slight body size, respectively; *P*-trend = 0.002] (**Table 4**). In analyses mutually adjusting for risk factors (parity or parous/

Table 3. Beta (β) coefficients and 95% CI for the associations between patients' clinicopathological characteristics and stromal microenvironment phenotypes (i.e., Ta-CTS and Ta-SCD) among patients participating in the GBHS.

		Ta-CTS				Ta-SCD			
		Partially adjusted		Multivariable		Partially adjusted		Multivariable	
Characteristic	Number	β (95% CI)	*P*	β (95% CI)	*P*	β (95% CI)	*P*	β (95% CI)	*P*
Age (years)[a]									
<35	185	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
35–44	183	−1.38 (−5.05, 2.29)		−1.60 (−5.19, 1.98)		−0.94 (−2.88, 1.00)		−0.69 (−2.64, 1.25)	
45–54	151	−1.59 (−5.19, 2.01)		−2.58 (−6.08, 0.93)		−1.42 (−3.33, 0.47)		−0.80 (−2.71, 1.11)	
≥55	167	−1.06 (−4.57, 2.45)	0.76	−1.44 (−4.85, 1.97)	0.62	0.04 (−1.81, 1.90)	0.51	0.33 (−1.52, 2.19)	0.31
ER status									
Negative	287	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Positive	307	3.20 (0.82, 5.57)	0.009	2.91 (0.19, 5.63)	0.03	−1.10 (−2.30, 0.11)	0.07	−0.60 (−2.39, 1.19)	0.49
PR status									
Negative	282	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Positive	309	0.13 (−3.04, 3.28)	0.93	−2.05 (−6.08, 0.93)	0.19	−0.18 (−1.65, 1.28)	0.80	0.38 (−1.60, 2.36)	0.69
HER2 status									
Negative	422	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Positive	137	3.58 (0.67, 6.49)	0.010	4.39 (1.43, 7.35)	0.005	0.26 (−1.30, 1.81)	0.74	−0.03 (−1.65, 1.60)	0.97
Subtype									
Luminal A	166	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Luminal B	188	−1.34 (−4.34, 1.65)	0.38	−0.60 (−3.54, 2.35)	0.69	−0.65 (−2.24, 0.94)	0.42	−0.78 (−2.37, 0.80)	0.33
HER2-enriched	47	2.59 (−2.07, 7.25)	0.28	3.88 (−0.67, 8.44)	0.09	1.39 (−1.09, 3.87)	0.27	0.99 (−1.46, 3.44)	0.43
TNBC	154	−4.33 (−7.58, −1.09)	0.009	−2.50 (−5.74, 0.73)	0.13	0.37 (−1.35, 2.61)	0.67	−0.36 (−2.10, 1.36)	0.68
Grade[a]									
Low	38	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Intermediate	168	−2.15 (−6.60, 2.30)		−2.07 (−6.70, 2.56)		3.69 (1.21, 6.18)		3.53 (1.03, 6.03)	
High	474	−9.41 (−13.48, −5.34)	<0.0001	−9.19 (−13.40, −4.98)	<0.0001	5.48 (3.09, 7.87)	<0.0001	5.22 (2.74, 7.71)	<0.0001
Size[a]									
<2 cm	23	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
2–5 cm	248	−2.32 (−8.66, 4.02)		−1.96 (−7.95, 4.03)		−0.01 (−3.19, 3.18)		0.07 (−3.10, 3.23)	
>5 cm	483	−1.00 (−7.04, 5.04)	0.45	−0.93 (−6.69, 4.83)	0.52	1.12 (−2.02, 4.26)	0.07	1.20 (−1.95, 4.36)	0.06

Partially adjusted linear regression models were adjusted for the relevant tumor characteristic, study site, and tissue area. Multivariable linear regression models were mutually adjusted for ER, PR, HER2, grade, and size in addition to age, study site, and tissue area. To estimate subtype-specific effect estimates, ER, PR, and HER2 were substituted by using the composite "Subtype" variable.

[a]*P* values for these variables were for trend tests. Missing covariate values on tumor characteristics were imputed using the MICE approach with appropriate variance adjustment by Rubin's Formula.

Table 4. Beta (β) coefficients and 95% CI for the associations between breast cancer risk factors and stromal microenvironment phenotypes (i.e., Ta-CTS and Ta-SCD) among patients participating in the GBHS.

		Ta-CTS				Ta-SCD			
		Partially adjusted		Multivariable		Univariable		Multivariable	
Characteristic	Number	β (95% CI)	*P*	β (95% CI)	*P*	β (95% CI)	*P*	β (95% CI)	*P*
Age at Menarche, years[a]									
<15	185	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
15	183	2.63 (−0.33, 5.58)		2.51 (−0.53, 5.55)		−0.75 (−2.31, 0.80)		−0.94 (−2.51, 0.62)	
16	151	0.72 (−2.38, 3.81)		0.52 (−2.65, 3.69)		−0.50 (−2.13, 1.14)		−0.39 (−2.02, 1.23)	
≥17	167	1.84 (−1.17, 4.85)	0.42	1.23 (−1.93, 4.39)	0.74	−1.75 (−3.34, −0.17)	0.04	−1.78 (−3.36, −0.19)	0.05
Parity									
Nulliparous	71	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Parous	719	0.80 (−2.71, 4.32)	0.65	1.08 (−2.55, 4.70)	0.56	2.45 (0.60, 4.31)	0.01	2.92 (1.04, 4.81)	0.002
Number of children[a]									
1	88	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
2	125	−1.08 (−4.55, 2.39)		−1.10 (−4.63, 2.43)		0.71 (−1.37, 2.80)		0.75 (−1.36, 2.87)	
3	136	0.08 (−3.34, 3.51)		0.05 (−3.42, 3.53)		−0.90 (−2.96, 1.16)		−0.84 (−2.93, 1.24)	
4	123	−0.14 (−3.62, 3.34)		−0.001 (−3.54, 3.54)		−0.03 (−2.12, 2.06)		−0.12 (−2.24, 2.00)	
≥5	247	−1.70 (−4.80, 1.39)	0.32	−1.26 (−4.50, 1.97)	0.58	0.56 (−1.30, 2.42)	0.63	0.57 (−1.37, 2.51)	0.73
Age at first birth (years)[a]									
<19	230	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
19–21	200	0.66 (−2.06, 3.38)		0.63 (−2.19, 3.45)		1.39 (−0.05, 2.83)		1.80 (0.34, 3.26)	
22–25	180	2.50 (−0.31, 5.31)		2.30 (−0.61, 5.22)		−0.37 (−1.86, 1.12)		0.05 (−1.46, 1.57)	
≥26	137	1.78 (−1.26, 4.82)		1.78 (−1.31, 4.87)	0.11	0.88 (−0.72, 2.50)	0.73	1.28 (−0.32, 2.89)	0.39
Breastfeeding, months[a]									
<13	240	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
13–18	415	0.60 (−1.99, 3.20)		0.73 (−1.94, 3.40)		0.61 (−0.76, 1.98)		0.75 (−0.63, 2.13)	
≥19	97	−2.08 (−5.72, 1.55)	0.41	−2.12 (−5.95, 1.71)	0.45	0.69 (−1.23, 2.61)	0.41	0.53 (−1.46, 2.52)	0.47
Parity/breastfeeding (months)									
Nulliparous	71	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Parous/<13	169	0.89 (−3.12, 4.89)	0.66	0.78 (−3.28, 4.83)	0.71	2.09 (−0.02, 4.21)	0.05	2.48 (0.37, 4.61)	0.02
Parous/13–18	415	1.50 (−2.18, 5.18)	0.42	1.63 (−2.11, 5.38)	0.39	2.76 (0.82, 4.70)	0.005	3.34 (1.39, 5.30)	0.001
Parous/≥19	97	−1.24 (−5.81, 3.32)	0.59	−1.27 (−5.91, 3.37)	0.59	2.73 (0.33, 5.13)	0.02	3.36 (0.93, 5.78)	0.007
Body Size[a]									
Slight	190	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Average	296	−0.74 (−3.40, 1.93)		−0.68 (−3.42, 2.05)		−0.41 (−1.80, 0.98)		−0.68 (−2.07, 0.72)	
Heavy	260	0.52 (−2.22, 3.28)	0.64	0.48 (−2.36, 3.33)	0.67	−1.82 (−3.22, −0.42)	0.008	−2.14 (−3.56, −0.72)	0.002
Family history									
None	726	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
Yes	55	0.62 (−3.33, 4.57)	0.76	0.89 (−3.10, 4.87)	0.66	2.31 (0.22, 4.40)	0.03	2.36 (0.28, 4.44)	0.02

Partially adjusted linear regression models were adjusted for age, study site, and tissue area. The overall multivariable linear regression model was mutually adjusted for age at menarche, parity (or parity/breastfeeding), body size, family history of breast cancer in a first degree relative in addition to age, study site, and tissue area. Number of children, age at first birth, and breastfeeding were included in a separate model restricted to parous women.

[a]*P* values for these variables were for trend tests.

Table 5. Beta coefficients (β) and 95% CI for the associations between breast cancer risk factors and Ta-SCD among patients with low- and high-grade breast cancer participating in the GBHS.

	Low-grade (grade 1)		Moderate–High grade (grade 2 and 3)		
	Multivariable		Multivariable		
Characteristic	**β (95% CI)**	***P***	**β (95% CI)**	***P***	***P*-het**
Parity					
Nulliparous	1.00 (reference)		1.00 (reference)		
Parous	−4.42 (-16.87, 8.04)	0.47	1.82 (−0.50, 4.15)	0.12	0.73
Parity/breastfeeding (months)					
Nulliparous	1.00 (reference)		1.00 (reference)		
Parous/<13	−8.96 (−23.96, 6.03)	0.22	1.41 (−1.13, 3.95)	0.27	
Parous/13–18	−6.30 (-21.55, 8.95)	0.39	1.90 (−0.50, 4.29)	0.12	
Parous/≥19	−14.97 (−36.93, 6.98)	0.17	2.08 (−0.83, 4.99)	0.16	0.92
Body Size					
Slight	1.00 (reference)		1.00 (reference)		
Moderate	5.41 (−3.70, 14.52)	0.47	−1.04 (−2.67, 0.59)	0.21	
Heavy	6.20 (−3.89, 16.30)	0.21	−2.62 (−4.22, 1.03)	0.001	0.04
Family history					
None	1.00 (reference)		1.00 (reference)		
Yes	9.61 (−4.30, 23.52)	0.16	1.39 (−0.95, 3.72)	0.24	0.16

Multivariable linear regression models were mutually adjusted for age at menarche, parity (or parity/breastfeeding), body size, family history of breast cancer in a first degree relative, age, study site, and tissue area. P-het: P-value for heterogeneity.

breastfeeding, body size, family history, age) and clinicopathological characteristics (ER, PR, HER2, grade, size), parity, body size, family history, and grade remained independently associated with Ta-SCD (Supplementary Table S2). We did not find any of the assessed risk factors to be associated with Ta-CTS (**Table 4**).

To investigate whether observed associations between risk factors and Ta-SCD were similarly captured by histologic grade, we tested the associations between risk factors and tumor grade. In analyses mutually adjusting for relevant risk factors (i.e., those associated with Ta-SCD—parity, body size, family history) and tumor clinicopathological characteristics (ER, PR, HER2, size, Ta-SCD), only ER and Ta-SCD were associated with higher grade tumors. Parity showed suggestive ($P = 0.08$) associations with moderate-to-high compared to low-grade disease (Supplementary Table S3). In head-to-head comparison of grade and Ta-SCD with respect to parity status (Supplementary Table S4), Ta-SCD ($P = 0.006$) and not grade ($P = 0.59$ and 0.18 for grades 2 and 3 vs. 1, respectively) was associated with parity after accounting for age, body size, family history as well as other clinicopathological characteristics (Supplementary Table S4).

In stratified analysis by age (Supplementary Table S5), the magnitude of the associations between body size and Ta-SCD were larger among younger than older women even though differences were not significant (*P*-heterogeneity = 0.20). On the other hand, the magnitude of associations with Ta-SCD were slightly larger among younger than older women for parity, parity/breastfeeding, and family history but differences were not significant. In analysis stratified by ER status, the magnitude of the observed associations between risk factors and Ta-SCD were larger among participants with ER^+ than ER^- disease even though differences were not significant. In particular, the magnitude of associations between parity (vs. nulliparous) and family history (vs. no family history) were ~2 and ~6-fold higher among ER^+ than ER^- patients, respectively (Supplementary Table S6). In analysis stratified by tumor grade (**Table 5**), we found the directionality of the associations of parity and body size with Ta-SCD to differ according to levels of histologic grade, however only body size related differences were significant (*P*-value for heterogeneity = 0.04). Whereas parity was inversely associated with Ta-SCD among patients with low-grade tumors, it was positively associated with Ta-SCD among patients with high-grade tumors (*P*-heterogeneity = 0.73). Increasing body size was positively associated with Ta-SCD among patients with low-grade tumors [β (95% CI) = 5.41 (−3.70, 14.52) and 6.20 (−3.89, 16.30) for average and heavy *vs.* slight body size, respectively; *P*-trend = 0.18], while among those with intermediate/high grade tumors increasing body size was inversely associated with Ta-SCD [β (95% CI) = −1.04 (−2.67, 0.59) and −2.62 (−4.22, −1.03) for average and heavy versus slight body size, respectively; *P*-trend = 0.001] (**Table 5**).

Discussion

In analyses of 790 patients with breast cancer from a population-based case-control study in Ghana, we utilized high-accuracy machine learning algorithms to characterize H&E-based stromal microenvironment phenotypes (i.e., Ta-CTS and Ta-SCD) and investigated relationships with breast cancer risk factors and tumor characteristics. We found several tumor characteristics to be associated with Ta-CTS and Ta-SCD. In general, patients with high-grade tumors tended to have lower Ta-CTS and higher Ta-SCD than those with low-grade disease. Several risk factors were associated with Ta-SCD, but not Ta-CTS in this population. In particular, parity, family history, and body size were associated with Ta-SCD, independently of other risk factors or relevant tumor characteristics. Our findings of associations between aggressive tumor characteristics and Ta-SCD, together with our observation that several risk factors were associated with Ta-SCD raise the possibility that epidemiological risk factors may impact the stromal microenvironment in tumors, with potential implications for tumor biology.

Parity is well-established as a protective factor against ER^+ breast cancers, with nulliparous women being several folds more likely to develop ER^+ disease than parous women (16). On the other hand, parity has been shown to promote the risk of aggressive, mostly triple negative, breast cancer particularly among younger women but risk persists even among older women (39). The precise mechanisms underpinning parity-related etiologic heterogeneity have yet to be defined. In the current study, we found Ta-SCD to be higher among parous than nulliparous women, which is consistent with experimental data showing that parity might influence breast tumor biology by inducing changes in the stromal microenvironment (40, 41). In particular, genes associated with immune, inflammation, and wound response pathways have consistently been shown to be significantly upregulated among parous than nulliparous women (14, 42). Higher Ta-SCD was associated with higher tumor grade in this and other populations (31). Accordingly, our parity and Ta-SCD related data add to the growing body of literature indicating that parity might be associated with a higher frequency of aggressive breast cancer phenotypes through stroma-related pathways (40). This notion is particularly relevant among women in Ghana (33) and other sub-Saharan African populations who tend to have higher parity in parallel with higher incidence rates of aggressive (mostly high grade) breast cancers (25, 27). Although breastfeeding is thought to attenuate the associations between parity and aggressive breast cancer (20), the association between parity and high Ta-SCD that we found persisted irrespective of breastfeeding duration. This finding could be explained in light of studies showing that post-lactational changes in mammary tissue, such as upregulation of genes related to immune response and development, do not revert to nulliparous levels several years after lactation (42).

Our findings of higher Ta-SCD levels in relation to positive family history and, to a lesser degree, earlier age at menarche are consistent with their well-established associations with elevated breast cancer risk (43). The precise mechanisms by which a positive family history of breast cancer increases breast cancer risk is yet to be fully defined but is closely linked to a higher probability of carrying pathogenic variants in breast cancer predisposition genes. Mutations in breast cancer predisposition genes such as *BRCA1*, *BRAC2*, and *TP53*, lead to uncontrolled proliferation via perturbations in DNA repair mechanisms or loss of cell cycle control (44, 45). Because these perturbations are not limited to epithelial cells, our family history and stromal microenvironment-related findings might suggest that the associations between family history and breast cancer incidence and/or tumor biology may be partly mediated through stromal changes. Recent data from a large-scale international study showed that protein-truncating variants and pathogenic missense mutations in *BRCA1*, *RAD51C*, *RAD51D*, and *BARD1* more strongly predisposed to TNBC than other breast cancer subtypes (46). However, this study comprised mostly women of European ancestry populations and the tissue mechanisms by which genetic perturbations influenced TNBC incidence was not studied. Data from other studies showed that differences in ancestry-specific immunologic landscapes correlated with differences in TNBC biology and clinical outcomes (47). Our data are consistent with the notion that stroma-related changes (which encompass immunologic response) might be a pathway by which familial factors may impact breast cancer incidence and tumor biology. Further studies are needed to characterize the relationships between specific germline pathogenetic variants and stromal microenvironment phenotypes among women from sub-Saharan African populations.

We found larger body size to be inversely associated with Ta-SCD, a finding that is not consistent with the notion that adiposity engenders a chronic inflammatory state in mammary tissues (48). Immune exhaustion might be one of several plausible explanations for the inverse association that we observed between larger body size and Ta-SCD in this population (49). In general, T cells mediate several aspects of tumor-associated inflammation and immune response, but they are themselves regulated by a complex crosstalk involving cancer cells, inflammatory cells, stromal cells, and cytokines. Persistent inflammation can overwhelm the T-cell response and trigger their terminal differentiation, leading to the emergence of "exhausted" T-cell phenotypes with diminished capacity to mount immune or inflammatory responses (49). Although speculative, the combination of obesity, a proinflammatory state, and prolonged and episodic exposure to conditions such as malaria may predispose to immune exhaustion among women in sub-Saharan Africa. Immune exhaustion may, in turn, pave way for the development of immunologically 'cold' but aggressive breast cancer subtypes via immunoediting (50). Immune exhaustion may also be the consequence of heightened immune response to high-grade disease, which is more common among obese women (21). In support of this idea, the inverse association of increasing body size and Ta-SCD that we found was limited to women with higher grade disease. Among those with low-grade disease, we found a positive association between increasing body size and increasing Ta-SCD, as would be expected. Because ~95% of the women in the current study had intermediate or high-grade tumors, our overall findings in relation to body size are likely reflective of the predominant tumor pathology in this population. These findings could also be explained by the well-recognized paradoxical effect of obesity to cause suppression or activation of inflammatory response pathways in breast cancer (51, 52). Further studies involving detailed exposure assessment, including past medical and infectious disease history, in conjunction with detailed molecular characterization of the stromal microenvironment, will be required to lend additional insights into the relationships between body size and stromal biology among women from sub-Saharan Africa.

Important strengths of this study include the relatively large sample size in an underrepresented population, detailed risk factor information, availability of standard H&E-stained sections, and the innovative application of digital pathology and high-accuracy machine learning algorithms to characterize stromal microenvironment phenotypes (Ta-CTS and Ta-SCD) using H&E-stained images. In terms of limitations, lack of follow-up data on clinical outcomes precluded our ability to investigate the prognostic relevance of Ta-CTS and Ta-SCD. Nevertheless, these stromal features demonstrated associations with other clinicopathologic characteristics similar to what has been reported in other populations (29, 31), suggesting that they may be similarly associated with clinical outcomes among patients from sub-Saharan Africa. Multiple testing is another potential limitation of this study. To address this limitation, we used Bonferroni correction, following which histologic grade remained significantly associated with Ta-CTS and Ta-SCD. Parity and body size remained significant following Bonferroni correction while family history was not significant at the Bonferroni threshold ($P = 0.003$).

In conclusion, results from our study indicate that host factors that impact the incidence of breast cancer in the general population, such as parity, body size, and family history, might also be associated with specific stromal microenvironment changes in invasive breast cancer tissues. The findings also raise the possibility that, although currently underappreciated, the stromal microenvironment may

hold important clues into the etiopathogenesis of aggressive breast cancers, which tend to predominate among women in sub-Saharan Africa. Further studies are warranted to understand whether the stromal microenvironment in premalignant tissues influences subsequent tumor biology and how host (genetic and non-genetic) factors may contribute to this process.

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

M. Abubakar: Conceptualization, resources, data curation, formal analysis, supervision, validation, investigation, methodology, writing–original draft, project administration, writing–review and editing. **T.U. Ahearn:** Resources, data curation, supervision, investigation, project administration, writing–review and editing. **M.A. Duggan:** Data curation, validation, methodology, writing–review and editing. **S. Lawrence:** Data curation, methodology, writing–review and editing. **E.K. Adjei:** Resources, data curation, writing–review and editing. **J.-N. Clegg-Lamptey:** Resources, data curation, writing–review and editing. **J. Yarney:** Resources, data curation, writing–review and editing. **B. Wiafe-Addai:** Resources, data curation, writing-review and editing. **B. Awuah:** Resources, data curation, writing–review and editing. **S. Wiafe:** Resources, data curation, writing–review and editing. **K. Nyarko:** Resources, data curation, writing–review and editing. **F.S. Aitpillah:** Resources, data curation, writing–review and editing. **D. Ansong:** Resources, data curation, writing–review and editing. **S.M. Hewitt:** Data curation, methodology, writing–review and editing. **L.A. Brinton:** Resources, data curation, funding acquisition, writing–review and editing. **J.D. Figueroa:** Resources, data curation, supervision, funding acquisition, methodology, writing–review and editing. **M. Garcia-Closas:** Conceptualization, resources, data curation, supervision, funding acquisition, methodology, project administration, writing–review and editing. **L. Edusei:** Conceptualization, resources, data curation, validation, methodology, writing–review and editing. **N. Titiloye:** Conceptualization, resources, data curation, validation, methodology, writing–review and editing.

Acknowledgments

We are grateful to all women who agreed to participate in the study and provided information and biospecimens. M. Abubakar, T.U. Ahearn, S. Lawrence, and J.D. Figueroa were supported by intramural research funds from the Division of Cancer Epidemiology and Genetics, NCI, NIH, USA. This research was supported by the Intramural Research Program of the NIH, NCI, Division of Cancer Epidemiology and Genetics, USA.

The Ghana Breast Health Study Team: Ghana Statistical Service, Accra, Ghana: Drs. Robertson Adjei and Lucy Afriyie. Korle Bu Teaching Hospital, Accra, Ghana: Dr. Anthony Adjei, Dr. Florence Dedey, Dr. Verna Vanderpuye, Victoria Okyne, Naomi Ohene Oti, Evelyn Tay, Dr. Adu-Aryee, Angela Kenu, and Obed Ekpedzor. Komfo Anoyke Teaching Hospital, Kumasi, Ghana: Marion Alcpaloo, Isaac Boakye, Bernard Arhin, Emmanuel Assimah, Samuel Ka-chungu, Dr. Joseph Oppong, and Dr. Ernest Osei-Bonsu. Peace and Love Hospital, Kumasi, Ghana: Prof Margaret Frempong, Emma Brew Abaidoo, Bridget Nortey Mensah, Samuel Amanama, Prince Agyapong, Debora Boateng, Ansong Thomas Agyei, Richard Opoku, and Kofi Owusu Gyimah. Memorial Sloan Kettering Cancer Center, NY, USA: Dr. Lisa Newman. National Cancer Institute, Bethesda, MD, USA: Maya Palakal and Jake Thistle. Westat, Inc.: Michelle Brotzman, Shelley Niwa, Usha Singh, and Ann Truelove. University of Ghana: Prof. Richard Biritwum.

Note

Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Received March 11, 2024; revised May 31, 2024; accepted July 1, 2024; published first July 3, 2024.

References

1. Rønnov-Jessen L, Bissell MJ. Breast cancer by proxy: can the microenvironment be both the cause and consequence? Trends Mol Med 2009;15:5–13.
2. Mao Y, Keller ET, Garfield DH, Shen K, Wang J. Stromal cells in tumor microenvironment and breast cancer. Cancer Metastasis Rev 2013;32:303–15.
3. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol 2012;196:395–406.
4. Soysal SD, Tzankov A, Muenst SE. Role of the tumor microenvironment in breast cancer. Pathobiology 2015;82:142–52.
5. Bombonati A, Sgroi DC. The molecular pathology of breast cancer progression. J Pathol 2011;223:307–17.
6. Cichon MA, Degnim AC, Visscher DW, Radisky DC. Microenvironmental influences that drive progression from benign breast disease to invasive breast cancer. J Mammary Gland Biol Neoplasia 2010;15:389–97.
7. Place AE, Huh SJ, Polyak K. The microenvironment in breast cancer progression: biology and implications for treatment. Breast Cancer Res 2011;13:227.
8. Dittmer J, Leyh B. The impact of tumor stroma on drug response in breast cancer. Semin Cancer Biol 2015;31:3–15.
9. Albini A, Sporn MB. The tumour microenvironment as a target for chemoprevention. Nat Rev Cancer 2007;7:139–47.
10. Degnim AC, Hoskin TL, Arshad M, Frost MH, Winham SJ, Brahmbhatt RA, et al. Alterations in the immune cell composition in premalignant breast tissue that precede breast cancer development. Clin Cancer Res 2017;23:3945–52.
11. Baumgarten SC, Frasor J. Minireview: inflammation: an instigator of more aggressive estrogen receptor (ER) positive breast cancers. Mol Endocrinol 2012;26:360–71.
12. Troester MA, Lee MH, Carter M, Fan C, Cowan DW, Perez ER, et al. Activation of host wound responses in breast cancer microenvironment. Clin Cancer Res 2009;15:7020–8.
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 Res Treat 2012;131:1003–12.
14. Rotunno M, Sun X, Figueroa J, Sherman ME, Garcia-Closas M, Meltzer P, et al. Parity-related molecular signatures and breast cancer subtypes by estrogen receptor status. Breast Cancer Res 2014;16:R74.
15. Martin DN, Boersma BJ, Yi M, Reimers M, Howe TM, Yfantis HG, et al. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. PLoS One 2009;4:e4531.
16. Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a review of the literature. Breast Cancer Res Treat 2014;144:1–10.
17. Anderson WF, Pfeiffer RM, Wohlfahrt J, Ejlertsen B, Jensen M-B, Kroman N. Associations of parity-related reproductive histories with ER± and HER2± receptor-specific breast cancer aetiology. Int J Epidemiol 2017;46:86–95.
18. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 2011;103:250–63.
19. Palmer JR, Boggs DA, Wise LA, Ambrosone CB, Adams-Campbell LL, Rosenberg L. Parity and lactation in relation to estrogen receptor negative breast cancer in African American women. Cancer Epidemiol Biomarkers Prev 2011;20:1883–91.
20. Palmer JR, Viscidi E, Troester MA, Hong C-C, Schedin P, Bethea TN, et al. Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER Consortium. J Natl Cancer Inst 2014;106:dju237.
21. Abubakar M, Chang-Claude J, Ali HR, Chatterjee N, Coulson P, Daley F, et al. Etiology of hormone receptor positive breast cancer differs by levels of histologic grade and proliferation. Int J Cancer 2018;143:746–57.
22. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor–defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev 2004;13:1558–68.
23. McCarthy AM, Friebel-Klingner T, Ehsan S, He W, Welch M, Chen J, et al. Relationship of established risk factors with breast cancer subtypes. Cancer Med 2021;10:6456–67.

24. Newman LA, Kaljee LM. Health disparities and triple-negative breast cancer in African American women: a review. JAMA Surg 2017;152:485–93.
25. Fregene A, Newman LA. Breast cancer in sub-Saharan Africa: how does it relate to breast cancer in African-American women? Cancer 2005;103: 1540–50.
26. Eng A, McCormack V, dos-Santos-Silva I. Receptor-defined subtypes of breast cancer in indigenous populations in Africa: a systematic review and meta-analysis. PLoS Med 2014;11:e1001720.
27. Jedy-Agba E, McCormack V, Adebamowo C, dos-Santos-Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. Lancet Glob Health 2016;4:e923–35.
28. Deshmukh SK, Srivastava SK, Tyagi N, Ahmad A, Singh AP, Ghadhban AAL, et al. Emerging evidence for the role of differential tumor microenvironment in breast cancer racial disparity: a closer look at the surroundings. Carcinogenesis 2017;38:757–65.
29. Downey CL, Simpkins SA, White J, Holliday DL, Jones JL, Jordan LB, et al. The prognostic significance of tumour–stroma ratio in oestrogen receptor-positive breast cancer. Br J Cancer 2014;110:1744–7.
30. de Kruijf EM, van Nes JGH, van de Velde CJH, Putter H, Smit VTHBM, Liefers GJ, et al. Tumor–stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. Breast Cancer Res Treat 2011;125:687–96.
31. Abubakar M, Zhang J, Ahearn TU, Koka H, Guo C, Lawrence SM, et al. Tumor-associated stromal cellular density as a predictor of recurrence and mortality in breast cancer: results from ethnically diverse study populations. Cancer Epidemiol Biomarkers Prev 2021;30:1397–407.
32. Brinton LA, Awuah B, Nat Clegg-Lamptey J, Wiafe-Addai B, Ansong D, Nyarko KM, et al. Design considerations for identifying breast cancer risk factors in a population-based study in Africa. Int J Cancer 2017;140:2667–77.
33. Figueroa JD, Davis Lynn BC, Edusei L, Titiloye N, Adjei E, Clegg-Lamptey J-N, et al. Reproductive factors and risk of breast cancer by tumor subtypes among Ghanaian women: a population-based case–control study. Int J Cancer 2020;147:1535–47.
34. Brinton L, Figueroa J, Adjei E, Ansong D, Biritwum R, Edusei L, et al. Factors contributing to delays in diagnosis of breast cancers in Ghana, West Africa. Breast Cancer Res Treat 2017;162:105–14.
35. Ahuno ST, Doebley A-L, Ahearn TU, Yarney J, Titiloye N, Hamel N, et al. Circulating tumor DNA is readily detectable among Ghanaian breast cancer patients supporting non-invasive cancer genomic studies in Africa. NPJ Precis Oncol 2021;5:83.
36. Abubakar M, Fan S, Bowles EA, Widemann L, Duggan MA, Pfeiffer RM, et al. Relation of quantitative histologic and radiologic breast tissue composition metrics with invasive breast cancer risk. JNCI Cancer Spectr 2021;5:pkab015.
37. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377–99.
38. Rubin DB. Multiple imputation for nonresponse in surveys. New York (NY): John Wiley and Sons; 1987.
39. Jung AY, Ahearn TU, Behrens S, Middha P, Bolla MK, Wang Q, et al. Distinct reproductive risk profiles for intrinsic-like breast cancer subtypes: pooled analysis of population-based studies. J Natl Cancer Inst 2022;114:1706–19.
40. McDaniel SM, Rumer KK, Biroc SL, Metz RP, Singh M, Porter W, et al. Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. Am J Pathol 2006;168:608–20.
41. McCready J, Arendt LM, Rudnick JA, Kuperwasser C. The contribution of dynamic stromal remodeling during mammary development to breast carcinogenesis. Breast Cancer Res 2010;12:205.
42. Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, Zhong H, Slifker M, Peri S, et al. Genomic signature of parity in the breast of premenopausal women. Breast Cancer Res 2019;21:46.
43. McPherson K, Steel CM, Dixon JM. ABC of breast diseases: breast cancer—epidemiology, risk factors, and genetics. BMJ 2000;321:624–8.
44. Stoppa-Lyonnet D. The biological effects and clinical implications of BRCA mutations: where do we go from here? Eur J Hum Genet 2016;24:S3–9.
45. Abubakar M, Guo C, Koka H, Sung H, Shao N, Guida J, et al. Clinicopathological and epidemiological significance of breast cancer subtype reclassification based on p53 immunohistochemical expression. NPJ Breast Cancer 2019;5:20.
46. Breast Cancer Association Consortium; Mavaddat N, Dorling L, Carvalho S, Allen J, González-Neira A, Keeman R, et al. Pathology of tumors associated with pathogenic germline variants in 9 breast cancer susceptibility genes. JAMA Oncol 2022;8:e216744.
47. Martini R, Delpe P, Chu TR, Arora K, Lord B, Verma A, et al. African ancestry–associated gene expression profiles in triple-negative breast cancer underlie altered tumor biology and clinical outcome in women of African descent. Cancer Discov 2022;12:2530–51.
48. Iyengar NM, Hudis CA, Dannenberg AJ. Obesity and inflammation: new insights into breast cancer development and progression. Am Soc Clin Oncol Educ Book 2013;33:46–51.
49. Mendes F, Domingues C, Rodrigues-Santos P, Abrantes AM, Gonçalves AC, Estrela J, et al. The role of immune system exhaustion on cancer cell escape and anti-tumor immune induction after irradiation. Biochim Biophys Acta 2016;1865:168–75.
50. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 2002;3:991–8.
51. Wang Z, Aguilar EG, Luna JI, Dunai C, Khuat LT, Le CT, et al. Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. Nat Med 2019;25:141–51.
52. Floris G, Richard F, Hamy A-S, Jongen L, Wildiers H, Ardui J, et al. Body mass index and tumor-infiltrating lymphocytes in triple-negative breast cancer. J Natl Cancer Inst 2021;113:146–53.