Am J Cancer Res 2024;14(11):5434-5445
www.ajcr.us /ISSN:2156-6976/ajcr0159713

Original Article

Reproductive factors and expression of stem cell markers in women with incident benign breast disease

Tengteng Wang[1,2,3], Yujing J Heng[4], Brian R Sardella[4], Bernard Rosner[5,6], Rulla M Tamimi[7], Lusine Yaghjyan[8]

[1]Section of Cancer Epidemiology and Health Outcomes, Rutgers Cancer Institute, New Brunswick, NJ, USA; [2]Division of Medical Oncology, Robert Wood Johnson Medical School, New Brunswick, NJ, USA; [3]Department of Biostatistics and Epidemiology, Rutgers School of Public Health, Pitscataway, NJ, USA; [4]Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; [5]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [6]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [7]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA; [8]Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, FL, USA

Received August 9, 2024; Accepted October 21, 2024; Epub November 15, 2024; Published November 30, 2024

Abstract: Reproductive factors are well-established risk factors for breast cancer. The prevailing hypothesis suggested that stem cell changes may be the key underlying mechanisms, but epidemiological evidence has been notably scarce. Herein we examined the relationship between reproductive risk factors and the expression of well-established stem cell markers CD44, CD24, and ALDH1A1 in benign breast biopsy non-cancerous samples. Our study included 735 participants from the Nurses' Health Study II who were diagnosed with biopsy-confirmed incident benign breast disease (BBD). Reproductive history and other BBD/breast cancer risk factors were measured from self-reported biennial questionnaires. Immunohistochemistry was performed on breast tissue microarrays from normal terminal ductal-lobular units (TDLU) cores. Marker expression in epithelium and stroma was quantified using semi-automated image analysis. The generalized linear regression was used to examine the associations of reproductive factors with the positive expression of CD44, CD24, and ALDH1A1, adjusted for known breast cancer risk factors. Age at first birth ≥30 years old (vs. <25 years) was associated with lower ALDH1A1 expression in the epithelium (β for ≥30 vs. <25 years = -0.30, 95% CI -0.57; -0.03, p-trend = 0.03). Parity, breastfeeding, age at menarche, and the time interval between menarche and age at first birth were not associated with the expression of any of the three markers in epithelium or stroma. These findings suggest age at first birth may influence the ALDH1A1 expression in breast tissue. Our study added to the very limited evidence regarding the potential impact of reproductive factors on breast stem cell markers.

Keywords: Benign breast disease, stem cell markers, breast cancer risk

Introduction

Reproductive factors, especially those related to childbearing, are well-established risk factors for breast cancer [1-4]. Several studies have demonstrated that greater parity, younger age at first birth, and longer duration of breastfeeding were associated with lower risk of developing breast cancer, with consistent findings pointing towards its role in modifying circulating hormone levels [5, 6], decreasing mammographic breast density [7-10], influencing epithelial-stroma interactions [11], and reducing the risk of benign breast disease (BBD) [12, 13]. Additionally, reproductive factors may also influence epigenetic modifications, gene expression profiles, and tissue composition in the breast [11, 14, 15], further affecting women's susceptibility to breast cancer.

Recently, another prevailing hypothesis suggested that stem cell pool reduction may be a key mechanism underlying the protective effect of these childbearing-related factors on breast cancer carcinogenesis [16]. According to the stem cell hypothesis, the likelihood of developing breast cancer may be linked to both the quantity and mitotic activity of the stem cell pool in the breast [16]. Additionally, within the mammary gland, only stem cells possess the

https://doi.org/10.62347/KLAC9365

unique ability to accumulate oncogenic changes throughout a woman's life [16]. Previous studies also indicate that pregnancy prompts differentiation in a subset of stem cells, thus lowering the pool of cells susceptible to transformation within the breast [16-20]. Furthermore, pregnancies occurring earlier in life, as opposed to later, may effectively reduce the quantity of mammary stem cells [16]. This finding aligns with observations that a younger age at first birth and a shorter interval between menarche and first birth are protective against breast cancer risk [17, 19, 20]. Moreover, pregnancy is believed to modulate signaling pathways integral to the function of mammary stem cells [16]. Consequently, the changes in the size and activity of the stem cell pool might elucidate the link between parity and breast cancer.

However, epidemiological studies exploring the associations between reproductive factors and breast stem cells have been notably scarce. We conducted a previous analysis within a cohort of 439 cancer-free women in the Nurses' Health Study (NHS) and NHSII, in which we found that age at menarche, age at first birth, and the time interval between these two events were inversely associated with expressions of stem cell biomarkers [21]. In the current study, we present a larger new cohort of women with incident BBD diagnosis within NHSII, to continue examining the relationship between various early-life reproductive factors and the expression of stem cell markers in benign breast biopsy samples. We focused on CD44, CD24, and ALDH1A1, given that they are well-established cancer stem cell markers [22]. We hypothesized that reproductive factors known to confer protective effects (e.g., earlier age of first birth and a higer parity) for breast cancer may be associated with lower stem cell markers' expression.

Materials and methods

Study population

Our study encompasses participants with incident biopsy-confirmed BBD from the NHSII. The NHSII consists of registered nurses across the U.S., aged 25-42 years at recruitment [23]. The initial 1989 NHSII questionnaire and all subsequent biennial questionnaires asked women to report whether they had ever received a physician diagnosis of fibrocystic or other BBD and whether the diagnosis was confirmed by biopsy or aspiration [24]. Cumulative response rates were >90%, and response rates to each questionnaire were very similar among women with and without previously reported BBD [25]. Updated biennial questionnaires provided follow-up data on breast cancer risk factors such as reproductive history, exogenous hormone usage, alcohol consumption, and disease diagnoses, including BBD, confirmed through medical records [26]. Details about the study design methods for the NHSII and incident BBD study have been published previously [24, 26-28].

Women who reported a first diagnosis of biopsy-confirmed BBD were contacted to confirm the diagnosis and to acquire permission to review their pathology specimens. After permission was granted, benign breast biopsy slides were collected from hospital pathology departments and were coded and submitted to the study pathologists [24]. Exclusions were made for in situ or invasive carcinoma or unknown lesion type at biopsy. Our current analysis includes 735 women with complete reproductive factor data and stem cell marker staining results.

The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. Consent was either explicitly given or implied by the questionnaire return.

Reproductive variables

Data on reproductive variables, including age at menarche, parity, age at first birth, age at last birth, and breastfeeding, were obtained from baseline and nearest biennial questionnaires preceding the biopsy date. Among eligible women with stem cell marker data (n = 735), nulliparous status and age at menarche data completion was 100%. In parous women with marker data (n = 634), the complete data on number of children, age at first birth, and breastfeeding were available for 100%, 98%, and 97% of women, respectively.

Age at first birth was categorized as <25, 25-29, ≥30 years and also modeled as a continuous variable. Parity was defined both as binary (nul-

liparous, parous) and categorical (1, 2, 3, ≥4 children), with the number of children among parous women also modeled as a continuous variable. Breastfeeding duration across all births was categorized as none to <1, 1- <12, 12- <24, and ≥24 months. Age at menarche was categorized as <12, 12, 13, >13 years and also modeled as a continuous variable. The interval between menarche and first birth and the time since last birth were treated as a continuous variable [21].

Benign breast biopsy confirmation and BBD subtypes

Hematoxylin and eosin (H&E)-stained breast tissue slides from women with BBD were independently reviewed by pathologists and blinded to the original BBD diagnosis [29, 30]. Questionable or confirmed atypia cases were jointly reviewed. Each BBD case was classified as non-proliferative, proliferative without atypia, or atypical hyperplasia (ductal or lobular) according to standard categorization [25, 31].

Tissue microarray (TMA) construction of BBD samples

After centralized review, formalin-fixed paraffin-embedded (FFPE) benign breast biopsy blocks were collected. H&E sections of corresponding tissue blocks were re-evaluated to identify proliferative lesions and normal terminal ductlobular units (TDLUs) for TMA core extraction. TMAs, constructed at the Dana Farber/Harvard Cancer Center (DF/HCC) Tissue Microarray Core Facility, contained up to 3 normal TDLU cores per woman. Our previously evaluated TMA construction method achieved a high success rate (76%) in capturing normal TDLUs [32].

Immunohistochemistry (IHC) for stem cell markers

The expression of the stem cell markers was evaluated by automated IHC techniques that allow the quantification of markers' expression levels and localization of the target signal to specific cells/structures. For each of the three markers, one 5-μm paraffin section was cut from a single TMA block and then stained with antibodies for CD44, CD24, and ALDH1A1 at the University of Florida Pathology Core Lab on DAKO AutostainerPlus according to the previously standardized protocol with commercial antibodies (DAKO AutostainerPlus, CD44 [DAKO] 1:25 dilution; CD24 [Invitrogen, Waltham, MA] 1:200 dilution and ALDH1A1 [Abcam, Cambridge, MA] 1:300 dilution). Details of this protocol have been described previously [21]. Briefly, slides were de-paraffinized with xylene and re-hydrated through decreasing concentrations of ethanol to water, including an intermediate step to quench endogenous peroxidase activity (3% hydrogen peroxide in methanol) and transferred to 1X TBS-T (Tris-buffered saline-Tween). For heat-induced antigen retrieval, sections were heated in a steamer while submerged in Citra (Biogenex, Fremont, CA) or Trilogy (Cell Marque, Rocklin, CA) for 30 minutes. Next, slides were 1) rinsed in 1XTBS-T and incubated with a universal protein blocker Sniper (Biocare Medical, Walnut Creek, CA) for 10 (for CD44 and ALDH1A1) or 15 minutes (for CD24); 2) rinsed in 1XTBS-T and co-incubated in primary antibody ALDH1A1, CD24, or CD44 for 1 hour; and 3) rinsed in 1XTBS-T followed by application of conjugated secondary antibody (Mach 2 goat anti-rabbit horse [or mouse] radish peroxidase-conjugated, Biocare Medical, Walnut Creek, CA) for 30 minutes. Detection of antibodies was achieved by incubating slides in 3'3' diaminobenzidine (Vector Laboratories Inc., Burlingame, CA) for 4 minutes. Slides were counterstained with hematoxylin (Biocare Medical, Walnut Creek, CA) 1:10 for 3 minutes and mounted with Cytoseal XYL (Richard-Allen Scientific, Kalamazoo, MI). The laboratory implemented standard quality control procedures.

Image analysis

Stained TMA sections were digitized at 40× using the PhenoImager HT (Akoya Biosciences, Marlborough, MA). QuPath v0.5.0 was used to quantify the immunoreactivity of the markers [33]. For each marker, we randomly selected 12 tissue cores of variable staining intensities on a TMA to train tissue segmentation into epithelium, stroma, and background. For each core, the expression level of each marker was evaluated on a continuous scale, representing the percentage (0-100%) of cells that stained positive (across all intensities) for a specific marker, assessed separately for epithelial and stromal cells [21, 34]. All analyses were limited to cores with at least 50 cells of specific tissue type (epithelium or stroma), consistent with

previously used approaches [35, 36]. We determined the minimum intensity to score a cell as immunoreactive [34]. We used a random forest object classifier and the "positive cell detection" command to optimize cell parameters and intensity thresholds for hematoxylin and cytoplasm (mean optical density) and used the default values for all other parameters. The focus was particularly on normal TDLU cores due to their relevance to early breast carcinogenesis and the heterogeneity and smaller number of benign lesion cores. The Spearman correlation between pathologist and computational assessments ranged between 0.40-0.64 for stroma and 0.66-0.68 for epithelium in normal TDLUs cores [35]. Staining results for stroma were available for 728, 714, and 714 women for CD44, CD24, and ALDH1A1, respectively; the staining results for epithelium were available for 725, 708, and 717 women, respectively.

Covariate information

Data on breast cancer risk factors were extracted from biennial questionnaires closest to the biopsy date. The postmenopausal status of participants was determined based on established criteria, considering factors like menstrual cycle cessation, oophorectomy, and hysterectomy [37, 38].

Statistical analysis

Multivariate-adjusted linear regression was utilized to explore the relationships between various reproductive factors and each marker's continuous expression (log-transformed) (separately for epithelium and stroma). We used a weighted average to summarize the marker expression level across available cores for a woman [35]. The risk estimates were adjusted for the following potential confounders: age (continuous), body mass index (BMI, continuous), a family history of breast cancer (yes/no), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia), and alcohol use (none, >0- <5, ≥5 g/day) were adjusted for the models. Additionally, in the analysis of the association of breastfeeding, the estimates were further adjusted for parity and age at first birth. The risk estimates were mutually adjusted for these two variables in analyzing the associations of parity and age at first birth. The estimates were adjusted for parity in the analysis for the interval between menarche and first birth.

The analyses of all reproductive variables except nulliparity and age at menarche were limited to parous women only. Parity, age at first birth, and age at menarche were modeled both as continuous and categorical, and breastfeeding was modeled as categorical. The lowest category for parity (1 child), age at first birth (less than 25 years), and breastfeeding (0 - less than 1 month) were used as the reference. To assess the overall trend for each categorical reproductive variable, we used respective medians within each category. The duration of the interval between menarche and first birth was modeled as a continuous variable. All statistical analyses were conducted using SAS software (version 9.4, SAS Institute, Cary, NC).

Results

In this study of 735 cancer-free women with incident BBD, 195 (26.5%) had non-proliferative disease, 480 (65.3%) had proliferative disease without atypia, and 60 (8.2%) had atypical hyperplasia. The average age at the biopsy was 44 years (range 27-63 years). Women in our sample were predominantly premenopausal at the biopsy (81.2%). The majority of women were parous (86.3%).

Table 1 presented the age-adjusted characteristics of 735 women with incident BBD according to parous status. The distribution of the three markers's expression across BBD subtypes in our study is presented in Supplementary Table 1.

In multivariate analysis (**Table 2**), age at first birth was inversely associated with ALDH1A1 expression in epithelium (β for 25-29 vs. <25 years = -0.17, 95% CI -0.39; 0.05; β for ≥30 vs. <25 years = -0.30, 95% CI -0.57; -0.03, p-trend = 0.03). We observed a suggestive positive association of parity with CD24 expression in epithelium which, however, did not reach statistical significance (p-trend = 0.08). Breastfeeding, age at menarche, the time interval between menarche and age at first birth, and time since last birth were not associated with

Table 1. Characteristics of women at the time of the biopsy, stratified by parous status

Characteristic	Parous n = 634	Nulliparous n = 101
Mean (SD)		
Age (years)	44.5 (6.0)	43.2 (6.2)
Age at menopause (years)	46.3 (4.9)	45.2 (9.7)
Body mass index (kg/m^2)	25.6 (5.6)	26.2 (7.2)
Alcohol use (grams/day)	3.8 (6.0)	5.1 (8.0)
Parity	2.3 (0.9)	NA
Age at first birth (years)	26.5 (4.5)	NA
Age at menarche (years)	12.6 (1.4)	12.4 (1.3)
CD44 normal TDLU Epithelium %	47.8 (24.6)	50.8 (21.1)
CD44 normal TDLU Stroma %	8.6 (8.9)	7.8 (7.1)
CD24 normal TDLU Epithelium %	73.3 (17.0)	71.7 (16.3)
CD24 normal TDLU Stroma %	43.5 (19.0)	39.7 (17.7)
ALDH1A1 normal TDLU Epithelium %	14.8 (12.1)	13.0 (9.4)
ALDH1A1 normal TDLU Stroma %	26.1 (20.9)	23.6 (17.8)
Percentages		
Breastfeeding		
0- <1 month	18	NA
1- <12 months	31	NA
12- <24 months	27	NA
≥24 months	24	NA
Family history of breast cancer	10	8
Smoking status		
Never smoked	67	56
Past smoker	24	29
Current smoker	9	15
Menopausal status/Postmenopausal hormone use		
Premenopausal	80	89
Never used	4	2
Past use	2	1
Current use	8	5
Benign breast disease subtypes		
Non-proliferative	66	29
Proliferative without atypia	66	62
Proliferative with atypia	8	10

Abbreviations: SD = standard deviation, NA = not applicable. Note: Values are means (SD) and percentages and are standardized to the age distribution of the study population.

the expression of any of the three markers in epithelium or stroma. The results were similar in reduced models adjusted for age and BMI (Supplementary Table 2).

Discussion

In our study of 735 cancer-free women, we explored the associations of childbearing-related reproductive factors with the expression of stem cell markers CD44, CD24, and ALDH1A1 in benign breast biopsy samples. We found significant inverse associations of age at first birth with ALDH1A1 expression in epithelium. No associations were observed for the remaining reproductive factors.

Our findings add to the limited evidence on the relationship between reproductive factors and stem cell marker expression in cancer-free women. Established theories suggest that reproductive events, such as pregnancy and lactation, induce the differentiation of mammary stem cells, which may reduce the pool of undifferentiated cells that can give rise to malignancies. Several potential underlying biological mechanisms are often tied to hormonal regulation [39], immune regulation [40], epigenetic reprogramming [15], and tissue remodeling [14], which can further affect stem cell activity in breast tissue. Our findings, which demonstrate differential expression of stem cell markers based on reproductive history, align with these theories by showing that reproductive events may indeed modify the stem cell landscape in ways that reduce breast cancer risk, particularly in parous women. However, our findings also present nuanced insights that challenge some aspects of these theories. Specifically, it has been suggested that full-term pregnancy may exert its protective effect on the risk of breast cancer [41] by decreasing the number of stem cells [4, 42-46]; however,

Table 2. Associations of reproductive factors with log-transformed expression of stem cell markers in benign breast biopsy samples (β coefficients and 95% Confidence intervals)

Reproductive factor	CD44				CD24				ALDH1A1			
	N	In Epithelium	N	In Stroma	N	In Epithelium	N	In Stroma	N	In Epithelium	N	In Stroma
Nulliparity[a]												
Nulliparous	98	0.15 (-0.04; 0.35)	100	0.02 (-0.37; 0.41)	95	-0.02 (-0.08; 0.05)	97	-0.08 (-0.20; 0.04)	97	-0.2×10^{-2} (-0.25; 0.24)	97	-0.18 (-0.62; 0.26)
Parous	627	Reference	628	Reference	613	Reference	617	Reference	620	Reference	617	Reference
Breastfeeding, months[b]												
0- <1	111	Reference	111	Reference		Reference	109	Reference	109	Reference	107	Reference
1- <12	185	0.01 (-0.22; 0.24)	185	-0.14 (-0.59; 0.31)	109	0.02 (-0.05; 0.09)	185	-0.08 (-0.22; 0.06)	186	0.15 (-0.13; 0.43)	185	0.13 (-0.34; 0.59)
12- <24	164	0.04 (-0.20; 0.28)	164	0.05 (-0.52; 0.43)	183	0.03 (-0.04; 0.10)	161	-0.04 (-0.19; 0.11)	159	0.31 (0.02; 0.60)	159	0.38 (-0.11; 0.88)
≥24	150	0.01 (-0.25; 0.26)	151	-0.13 (-0.64; 0.37)	160	-0.01 (-0.09; 0.07)	145	-0.05 (-0.20; 0.11)	149	0.24 (-0.07; 0.55)	149	0.24 (-0.29; 0.76)
p-trend		0.93		0.79	144	0.81		0.93		0.14		0.32
Parity[c]												
1	107	Reference	107	Reference	104	Reference	104	Reference	105	Reference	103	Reference
2	277	-0.21 (-0.43; 0.02)	277	-0.38 (-0.82; 0.07)	271	-0.01 (-0.08; 0.06)	272	-0.05 (-0.19; 0.09)	273	-0.08 (-0.36; 0.19)	273	-0.51 (-0.97; -0.04)
3	169	-0.16 (-0.40; 0.09)	170	-0.33 (-0.82; 0.15)	166	0.03 (-0.04; 0.11)	169	0.02 (-0.13; 0.17)	168	0.03 (-0.27; 0.33)	168	-0.08 (-0.58; 0.43)
≥4	63	-0.08 (-0.40; 0.23)	63	-0.15 (-0.77; 0.48)	62	0.06 (-0.03; 0.16)	62	0.08 (-0.11; 0.28)	63	-0.05 (-0.44; 0.34)	62	-0.05 (-0.69; 0.60)
p-trend		0.75		0.66		0.08		0.24		0.79		0.42
Parity continuous[c]	616	-0.01 (-0.09; 0.08)	617	-0.01 (-0.18; 0.16)	603	0.02 (-0.01; 0.04)	607	0.02 (-0.03; 0.07)	609	0.01 (-0.09; 0.12)	606	0.07 (-0.11; 0.24)
Age at 1st birth[d]												
<25	214	Reference	214	Reference	209	Reference	209	Reference	211	Reference	210	Reference
25-29	256	-0.08 (-0.25; 0.10)	256	-0.24 (-0.58; 0.11)	250	-0.001 (-0.06; 0.05)	253	-0.10 (-0.20; 0.01)	254	-0.17 (-0.39; 0.05)	253	-0.32 (-0.68; 0.04)
≥30	146	-0.01 (-0.23; 0.21)	147	-0.22 (-0.65; 0.21)	144	-0.01 (-0.07; 0.06)	145	-0.10 (-0.23; 0.03)	144	-0.30 (-0.57; -0.03)	143	-0.22 (-0.66; 0.23)
p-trend		0.87		0.27		0.87		0.12		0.03		0.28
Age at 1st birth, continuous (per 5 years)[d]	616	0.2×10^{-2} (-0.09; 0.09)	617	-0.03 (-0.21; 0.15)	603	-0.01 (-0.04; 0.02)	607	-0.04 (-0.09; 0.02)	609	-0.10 (-0.22; 0.01)	606	-0.04 (-0.23; 0.15)
Age at menarche[e]												
<12	154	0.08 (-0.13; 0.29)	157	-0.3×10^{-2} (-0.42; 0.42)	148	0.01 (-0.06; 0.07)	153	0.01 (-0.12; 0.14)	152	0.08 (-0.18; 0.34)	150	0.18 (-0.29; 0.65)
12	210	0.06 (-0.13; 0.26)	210	0.28 (-0.11; 0.67)	208	-0.02 (-0.08; 0.04)	209	-0.04 (-0.16; 0.08)	210	-0.4×10^{-2} (-0.25; 0.24)	209	0.01 (-0.42; 0.45)
13	213	0.07 (-0.12; 0.27)	213	0.25 (-0.14; 0.64)	205	-0.01 (-0.07; 0.05)	205	0.02 (-0.10; 0.14)	209	-0.5×10^{-2} (-0.29; 0.19)	210	-0.04 (-0.47; 0.40)
>13	148	Reference	148	Reference	147	Reference	147	Reference	146	Reference	145	Reference
p-trend		0.50		0.99		0.93		0.75		0.48		0.42

Age at menarche, continuous (per 5 years)[e]	725	-0.17 (-0.41; 0.08)	728	-0.09 (-0.58; 0.40)	708	-0.01 (-0.09; 0.07)	714	0.01 (-0.15; 0.16)	717	-0.13 (-0.43; 0.18)	714	-0.27 (-0.82; 0.28)
Time between menarche and age at 1st birth, continuous (per 5 years)[f]	616	0.02 (-0.07; 0.11)	617	-0.02 (-0.20; 0.15)	603	-0.01 (-0.03; 0.02)	607	-0.03 (-0.09; 0.02)	609	-0.08 (-0.18; 0.03)	606	-0.4×10^{-2} (-0.18; 0.17)
Time since last birth, continuous (per 5 years)[b]	616	-0.02×10^{-2} (-0.17; 0.16)	617	0.19 (-0.14; 0.51)	603	-0.5×10^{-2} (-0.06; 0.05)	607	-0.07 (-0.18; 0.03)	609	-0.19 (-0.39; 0.01)	606	-0.25 (-0.58; 0.09)

[a]Adjusted for age (continuous), BMI (continuous), age at menarche (<12, 12, 13, >13), a family history of breast cancer (Yes/No), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia),and alcohol use (none, >0- <5, ≥5 g/day). [b]Among parous women only: adjusted for age (continuous), BMI (continuous), age at menarche (<12, 12, 13, >13), parity, age at first child's birth, a family history of breast cancer (Yes/No), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia), and alcohol use (none, >0- <5, ≥5 g/day). [c]Among parous women only: adjusted for age (continuous), BMI (continuous), age at first birth, age at menarche (<12, 12, 13, >13), a family history of breast cancer (Yes/No), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia), and alcohol use (none, >0- <5, ≥5 g/day). [d]Among parous women only: adjusted for age (continuous), BMI (continuous), parity, age at menarche (<12, 12, 13, >13), a family history of breast cancer (Yes/No), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia), and alcohol use (none, >0- <5, ≥5 g/day). [e]Adjusted for age (continuous), BMI (continuous), parous status (nulliparous, parous), a family history of breast cancer (Yes/No), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia), and alcohol use (none, >0- <5, ≥5 g/day). [f]Among parous women only: adjusted for age (continuous), BMI (continuous), parity, a family history of breast cancer (Yes/No), menopausal status/postmenopausal hormone use (premenopausal, postmenopausal/no hormones, postmenopausal/past hormones, postmenopausal/current hormones, postmenopausal/unknown hormone use status), benign breast disease subtype (non-proliferative, proliferative without atypia, proliferative with atypia), and alcohol use (none, >0- <5, ≥5 g/day).

we did not find significant associations of parous status with the expression of any of the three markers, though nulliparous women had suggestive evidence of greater expression of CD44 in both epithelium and stroma. A previous investigation exploring potential links between parity and stem cell population size by analyzing the proportion of ALDH1A1 positive cells in breast epithelium also reported no significant differences across nulliparous and parous women [47]. Our recent analysis [21] conducted among 439 women with BBD in NHS and NHSII nested case-control study of breast cancer also found no associations of parity with any of the three markers.

In our current study, we noticed suggestive evidence of a positive association between the number of children and epithelial CD24 expression. Previous studies suggest that combined expression of CD44(+), CD24(-/low), and ALDH1(high) expression could be used to characterize two largely non-overlapping populations of breast cancer stem cells that have epithelial-like and mesenchymal-like phenotypes, respectively [48-50]. Even though we were unable to assess the combined expression of these markers on a cell-by-cell basis in our study, our findings of positive associations of parity with CD24 provide indirect evidence that the observed associations may be reflective of a reduction in the number and/or activity of breast stem cells.

We observed a significant inverse association between the age at first birth and ALDH1A1 expression in epithelial cells. These results are consistent with our previous findings from a smaller nested case-control study in NHS/NHSII cohorts with 439 women [21]. However, they contradict our previous hypothesis that earlier age of first birth may help reduce the size and activity of the stem cell pool in the breast [16, 51]. These inconsistent findings across animal and observational studies highlight the complexity of reproductive history's impact on breast cancer risk and suggest that not all reproductive events confer protective effects in the same manner. There is a crucial need to further elucidate the biological changes of stem cell behavior occurring during pregnancy and to explore how stem cells interact with other key breast cellular components of the tissue microenvironment [41]. Specifically, we have previously shown associations of stem cells with mammographic breast density [52-54], a strong breast cancer risk factor that has been linked both to reproductive factors [10, 55, 56] as well as pro-inflammatory tissue microenvironment [57, 58]. Further research is needed to build on our work by studying the interaction between reproductive history, stem cell markers, and local tissue microenvironment in order to better understand underlying biological pathways behind these complex associations.

Our study is the second and the largest to date to explore associations of several key reproductive risk factors with the expression of breast stem cell markers CD44, CD24, and ALDH1A1 in cancer-free women with BBD biopsy samples. We utilized data from the NHSII, an established cohort with more than 30 years of follow-up, with confirmation of BBD status and comprehensive breast cancer risk factor information to be adjusted for. There are also a few limitations. First, while the data from NHSII were gathered in a prospective manner, the possibility of measurement inaccuracies for certain reproductive factors, especially among postmenopausal women, remains possible. Second, it's important to note that our biopsy samples were obtained from the nomal TDLUs of the breast tissue. However, our previous studies have indicated that this tissue sampling method still yields robust evidence for measuring breast tissue involution [59], for identification of breast cancer biomarkers [60-62], and for establishing the correlation with various breast cancer risk factors [63]. Third, our study includes only cancer-free women who underwent routine clinical biopsies for BBD diagnosis potentially limiting their generalizability; however, given that our analysis focuses on normal TDLUs, we believe the results could still apply to a broader spectrum of women without BBD rather than only to women with BBD [11]. Furthermore, in our study, we did not employ co-localization techniques for the IHC staining, which precluded us from evaluating the combined expression of these markers at the individual cell level [64]. Additionally, all our BBD diagnoses were confirmed by biopsy or aspiration. We acknowledge that breast surgical procedures may modulate cancer biology [65-68], but this may result in non-differential impact on

marker status in our study given we used non-cancerous tissues. Finally, since the majority of our study participants were premenopausal, we were unable to conduct stratified analyses by menopausal status.

In conclusion, our study adds to the very limited evidence on how childbearing-related reproductive factors might influence the expression of key breast stem cell markers, potentially impacting breast cancer risk. The observed inverse associations of age at first birth with ALDH1A1 expression, along with the potential influence of parity on CD24 expression, underscore the need for further research to unravel the complex underlying biological mechanisms. Given that BBD is a known risk factor for breast cancer, our results might inform early detection strategies, particularly for women with specific reproductive histories. Our findings also shed light on the importance of integrating biomarkers, such as stem cell markers, in screening protocols to identify women at higher risk for progression to malignancy. Future studies are warranted to confirm these findings and advance our understanding of the role of stem cells in breast cancer etiology.

Acknowledgements

The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. The authors also would like to thank Zongmin Liu for his assistance in editing the first draft manuscript and downloading EndNote citation files. This work was supported by the National Cancer Institute at the National Institutes of Health [R01 CA240341 to Lusine Yaghjyan, CA131332, CA175080, P01 CA087969 to Rulla Tamimi, K99 CA 267557 and R00 CA 267557 to Tengteng Wang, and U01 CA176726 to Walter Willett], Avon Foundation for Women, Susan G. Komen for the Cure®, and Breast Cancer Research Foundation.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Lusine Yaghjyan, Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, FL, USA. E-mail: lyaghjyan@ufl.edu

References

[1] Woolcott CG, Koga K, Conroy SM, Byrne C, Nagata C, Ursin G, Vachon CM, Yaffe MJ, Pagano I and Maskarinec G. Mammographic density, parity and age at first birth, and risk of breast cancer: an analysis of four case-control studies. Breast Cancer Res Treat 2012; 132: 1163-1171.

[2] Butt S, Borgquist S, Anagnostaki L, Landberg G and Manjer J. Parity and age at first childbirth in relation to the risk of different breast cancer subgroups. Int J Cancer 2009; 125: 1926-1934.

[3] Ramon JM, Escriba JM, Casas I, Benet J, Iglesias C, Gavalda L, Torras G and Oromi J. Age at first full-term pregnancy, lactation and parity and risk of breast cancer: a case-control study in Spain. Eur J Epidemiol 1996; 12: 449-453.

[4] Britt K, Ashworth A and Smalley M. Pregnancy and the risk of breast cancer. Endocr Relat Cancer 2007; 14: 907-933.

[5] Eliassen AH, Missmer SA, Tworoger SS, Spiegelman D, Barbieri RL, Dowsett M and Hankinson SE. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. J Natl Cancer Inst 2006; 98: 1406-1415.

[6] Fortner RT, Eliassen AH, Spiegelman D, Willett WC, Barbieri RL and Hankinson SE. Premenopausal endogenous steroid hormones and breast cancer risk: results from the nurses' health study II. Breast Cancer Res 2013; 15: R19.

[7] Modugno F, Ngo DL, Allen GO, Kuller LH, Ness RB, Vogel VG, Costantino JP and Cauley JA. Breast cancer risk factors and mammographic breast density in women over age 70. Breast Cancer Res Treat 2006; 97: 157-166.

[8] Yaghjyan L, Mahoney MC, Succop P, Wones R, Buckholz J and Pinney SM. Relationship between breast cancer risk factors and mammographic breast density in the Fernald community cohort. Br J Cancer 2012; 106: 996-1003.

[9] Vachon CM, Kuni CC, Anderson K, Anderson VE and Sellers TA. Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States). Cancer Causes Control 2000; 11: 653-662.

[10] Yaghjyan L, Colditz GA, Rosner B, Bertrand KA and Tamimi RM. Reproductive factors related to childbearing and mammographic breast

density. Breast Cancer Res Treat 2016; 158: 351-359.
[11] Yaghjyan L, Austin-Datta RJ, Oh H, Heng YJ, Vellal AD, Sirinukunwattana K, Baker GM, Collins LC, Murthy D, Rosner B and Tamimi RM. Associations of reproductive breast cancer risk factors with breast tissue composition. Breast Cancer Res 2021; 23: 70.
[12] Li J, Humphreys K, Ho PJ, Eriksson M, Darai-Ramqvist E, Lindström LS, Hall P and Czene K. Family history, reproductive, and lifestyle risk factors for fibroadenoma and breast cancer. JNCI Cancer Spectr 2018; 2: pky051.
[13] Johansson A, Christakou AE, Iftimi A, Eriksson M, Tapia J, Skoog L, Benz CC, Rodriguez-Wallberg KA, Hall P, Czene K and Lindström LS. Characterization of benign breast diseases and association with age, hormonal factors, and family history of breast cancer among women in Sweden. JAMA Netw Open 2021; 4: e2114716.
[14] Lu P, Weaver VM and Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol 2012; 196: 395-406.
[15] Dos Santos CO, Dolzhenko E, Hodges E, Smith AD and Hannon GJ. An epigenetic memory of pregnancy in the mouse mammary gland. Cell Rep 2015; 11: 1102-1109.
[16] Dall G, Risbridger G and Britt K. Mammary stem cells and parity-induced breast cancer protection- new insights. J Steroid Biochem Mol Biol 2017; 170: 54-60.
[17] Li CI, Malone KE, Daling JR, Potter JD, Bernstein L, Marchbanks PA, Strom BL, Simon MS, Press MF, Ursin G, Burkman RT, Folger SG, Norman S, McDonald JA and Spirtas R. Timing of menarche and first full-term birth in relation to breast cancer risk. Am J Epidemiol 2008; 167: 230-239.
[18] Ritte R, Tikk K, Lukanova A, Tjønneland A, Olsen A, Overvad K, Dossus L, Fournier A, Clavel-Chapelon F, Grote V, Boeing H, Aleksandrova K, Trichopoulou A, Lagiou P, Trichopoulos D, Palli D, Berrino F, Mattiello A, Tumino R, Sacerdote C, Quirós JR, Buckland G, Molina-Montes E, Chirlaque MD, Ardanaz E, Amiano P, Bueno-de-Mesquita HB, van Gils CH, Peeters PH, Wareham N, Khaw KT, Key TJ, Travis RC, Weiderpass E, Dumeaux V, Lund E, Sund M, Andersson A, Romieu I, Rinaldi S, Vineis P, Merritt MA, Riboli E and Kaaks R. Reproductive factors and risk of hormone receptor positive and negative breast cancer: a cohort study. BMC Cancer 2013; 13: 584.
[19] Rosner B and Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst 1996; 88: 359-364.
[20] Pike MC, Krailo MD, Henderson BE, Casagrande JT and Hoel DG. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature 1983; 303: 767-770.
[21] Yaghjyan L, Heng YJ, Baker GM, Bret-Mounet VC, Murthy D, Mahoney MB, Rosner B and Tamimi RM. Associations of reproductive breast cancer risk factors with expression of stem cell markers in benign breast tissue. Front Oncol 2024; 14: 1354094.
[22] Li W, Ma H, Zhang J, Zhu L, Wang C and Yang Y. Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. Sci Rep 2017; 7: 13856.
[23] Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE and Chavarro JE. Origin, methods, and evolution of the three nurses' health studies. Am J Public Health 2016; 106: 1573-1581.
[24] Su X, Colditz GA, Willett WC, Collins LC, Schnitt SJ, Connolly JL, Pollak MN, Rosner B and Tamimi RM. Genetic variation and circulating levels of IGF-I and IGFBP-3 in relation to risk of proliferative benign breast disease. Int J Cancer 2010; 126: 180-190.
[25] Tamimi RM, Rosner B and Colditz GA. Evaluation of a breast cancer risk prediction model expanded to include category of prior benign breast disease lesion. Cancer 2010; 116: 4944-4953.
[26] Farland LV, Tamimi RM, Eliassen AH, Spiegelman D, Collins LC, Schnitt SJ and Missmer SA. A prospective study of endometriosis and risk of benign breast disease. Breast Cancer Res Treat 2016; 159: 545-552.
[27] Baer HJ, Schnitt SJ, Connolly JL, Byrne C, Willett WC, Rosner B and Colditz GA. Early life factors and incidence of proliferative benign breast disease. Cancer Epidemiol Biomarkers Prev 2005; 14: 2889-2897.
[28] Jung MM, Colditz GA, Collins LC, Schnitt SJ, Connolly JL and Tamimi RM. Lifetime physical activity and the incidence of proliferative benign breast disease. Cancer Causes Control 2011; 22: 1297-1305.
[29] Jacobs TW, Byrne C, Colditz G, Connolly JL and Schnitt SJ. Pathologic features of breast cancers in women with previous benign breast disease. Am J Clin Pathol 2001; 115: 362-369.
[30] Tamimi RM, Byrne C, Baer HJ, Rosner B, Schnitt SJ, Connolly JL and Colditz GA. Benign breast disease, recent alcohol consumption, and risk of breast cancer: a nested case-control study. Breast Cancer Res 2005; 7: R555-562.
[31] Page DL, Dupont WD, Rogers LW and Rados MS. Atypical hyperplastic lesions of the female

breast. A long-term follow-up study. Cancer 1985; 55: 2698-2708.
[32] Collins LC, Wang Y, Connolly JL, Baer HJ, Hu R, Schnitt SJ, Colditz GA and Tamimi RM. Potential role of tissue microarrays for the study of biomarker expression in benign breast disease and normal breast tissue. Appl Immunohistochem Mol Morphol 2009; 17: 438-441.
[33] Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M and Hamilton PW. QuPath: open source software for digital pathology image analysis. Sci Rep 2017; 7: 16878.
[34] Baker GM, Bret-Mounet VC, Wang T, Veta M, Zheng H, Collins LC, Eliassen AH, Tamimi RM and Heng YJ. Immunohistochemistry scoring of breast tumor tissue microarrays: a comparison study across three software applications. J Pathol Inform 2022; 13: 100118.
[35] Yaghjyan L, Heng YJ, Baker GM, Bret-Mounet V, Murthy D, Mahoney MB, Mu Y, Rosner B and Tamimi RM. Reliability of CD44, CD24, and ALDH1A1 immunohistochemical staining: pathologist assessment compared to quantitative image analysis. Front Med (Lausanne) 2022; 9: 1040061.
[36] Turashvili G, Leung S, Turbin D, Montgomery K, Gilks B, West R, Carrier M, Huntsman D and Aparicio S. Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis. BMC Cancer 2009; 9: 165.
[37] Willett W, Stampfer MJ, Bain C, Lipnick R, Speizer FE, Rosner B, Cramer D and Hennekens CH. Cigarette smoking, relative weight, and menopause. Am J Epidemiol 1983; 117: 651-658.
[38] Stampfer MJ, Willett WC, Colditz GA, Rosner B, Speizer FE and Hennekens CH. A prospective study of postmenopausal estrogen therapy and coronary heart disease. N Engl J Med 1985; 313: 1044-1049.
[39] Asselin-Labat ML, Vaillant F, Sheridan JM, Pal B, Wu D, Simpson ER, Yasuda H, Smyth GK, Martin TJ, Lindeman GJ and Visvader JE. Control of mammary stem cell function by steroid hormone signalling. Nature 2010; 465: 798-802.
[40] Karin M and Greten FR. NF-kappaB: linking inflammation and immunity to cancer development and progression. Nat Rev Immunol 2005; 5: 749-759.
[41] Meier-Abt F, Bentires-Alj M and Rochlitz C. Breast cancer prevention: lessons to be learned from mechanisms of early pregnancy-mediated breast cancer protection. Cancer Res 2015; 75: 803-807.
[42] Murrow LM, Weber RJ, Caruso J, McGinnis CS, Gascard P, Borowsky AD, Desai TA, Thomson M, Tlsty T and Gartner ZJ. Pregnancy and obesity modify the epithelial composition and hormone signaling state of the human breast. bioRxiv 2020; 430611.
[43] Ghosh S, Gu F, Wang CM, Lin CL, Liu J, Wang H, Ravdin P, Hu Y, Huang TH and Li R. Genome-wide DNA methylation profiling reveals parity-associated hypermethylation of FOXA1. Breast Cancer Res Treat 2014; 147: 653-659.
[44] Belitskaya-Lévy I, Zeleniuch-Jacquotte A, Russo J, Russo IH, Bordás P, Åhman J, Afanasyeva Y, Johansson R, Lenner P, Li X, de Cicco RL, Peri S, Ross E, Russo PA, Santucci-Pereira J, Sheriff FS, Slifker M, Hallmans G, Toniolo P and Arslan AA. Characterization of a genomic signature of pregnancy identified in the breast. Cancer Prev Res (Phila) 2011; 4: 1457-1464.
[45] Peri S, de Cicco RL, Santucci-Pereira J, Slifker M, Ross EA, Russo IH, Russo PA, Arslan AA, Belitskaya-Lévy I, Zeleniuch-Jacquotte A, Bordas P, Lenner P, Åhman J, Afanasyeva Y, Johansson R, Sheriff F, Hallmans G, Toniolo P and Russo J. Defining the genomic signature of the parous breast. BMC Med Genomics 2012; 5: 46.
[46] Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, Zhong H, Slifker M, Peri S, Ross EA, López de Cicco R, Zhai Y, Nguyen T, Sheriff F, Russo IH, Su Y, Arslan AA, Bordas P, Lenner P, Åhman J, Landström Eriksson AS, Johansson R, Hallmans G, Toniolo P and Russo J. Genomic signature of parity in the breast of premenopausal women. Breast Cancer Res 2019; 21: 46.
[47] Honeth G, Schiavinotto T, Vaggi F, Marlow R, Kanno T, Shinomiya I, Lombardi S, Buchupalli B, Graham R, Gazinska P, Ramalingam V, Burchell J, Purushotham AD, Pinder SE, Csikasz-Nagy A and Dontu G. Models of breast morphogenesis based on localization of stem cells in the developing mammary lobule. Stem Cell Reports 2015; 4: 699-711.
[48] Colacino JA, Azizi E, Brooks MD, Harouaka R, Fouladdel S, McDermott SP, Lee M, Hill D, Madden J, Boerner J, Cote ML, Sartor MA, Rozek LS and Wicha MS. Heterogeneity of human breast stem and progenitor cells as revealed by transcriptional profiling. Stem Cell Reports 2018; 10: 1596-1609.
[49] Ricardo S, Vieira AF, Gerhard R, Leitão D, Pinto R, Cameselle-Teijeiro JF, Milanezi F, Schmitt F and Paredes J. Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. J Clin Pathol 2011; 64: 937-946.

[50] Escudero Mendez L, Srinivasan M, Hamouda RK, Ambedkar B, Arzoun H, Sahib I, Fondeur J and Mohammed L. Evaluation of CD44+/CD24- and aldehyde dehydrogenase enzyme markers in cancer stem cells as prognostic indicators for triple-negative breast cancer. Cureus 2022; 14: e28056.

[51] Biro FM and Deardorff J. Identifying opportunities for cancer prevention during preadolescence and adolescence: puberty as a window of susceptibility. J Adolesc Health 2013; 52 Suppl: S15-20.

[52] Yaghjyan L, Esnakula AK, Scott CG, Wijayabahu AT, Jensen MR and Vachon CM. Associations of mammographic breast density with breast stem cell marker-defined breast cancer subtypes. Cancer Causes Control 2019; 30: 1103-1111.

[53] Yaghjyan L, Stoll E, Ghosh K, Scott CG, Jensen MR, Brandt KR, Visscher D and Vachon CM. Tissue-based associations of mammographic breast density with breast stem cell markers. Breast Cancer Res 2017; 19: 100.

[54] Yaghjyan L, Heng YJ, Baker GM, Murthy D, Mahoney MB, Rosner B and Tamimi RM. Associations of stem cell markers CD44, CD24 and ALDH1A1 with mammographic breast density in women with benign breast biopsies. Br J Cancer 2024; 131: 325-333.

[55] O'Driscoll J, Burton A, Maskarinec G, Perez-Gomez B, Vachon C, Miao H, Lajous M, López-Ridaura R, Eliassen AH, Pereira A, Garmendia ML, Tamimi RM, Bertrand K, Kwong A, Ursin G, Lee E, Qureshi SA, Ma H, Vinnicombe S, Moss S, Allen S, Ndumia R, Vinayak S, Teo SH, Mariapun S, Fadzli F, Peplonska B, Nagata C, Stone J, Hopper JL, Giles G, Ozmen V, Aribal ME, Schüz J, Van Gils CH, Wanders JOP, Sirous R, Sirous M, Hipwell J, Kim J, Lee JW, Hartman M, Li J, Scott C, Chiarelli AM, Linton L, Pollan M, Flugelman AA, Salem D, Kamal R, Boyd N, dos-Santos-Silva I, McCormack V and Mullooly M. Reproductive factors and mammographic density within the International Consortium of Mammographic Density: a cross-sectional study. Breast Cancer Res 2024; 26: 139.

[56] Yaghjyan L, Wang Z, Warner ET, Rosner B, Heine J and Tamimi RM. Reproductive factors related to childbearing and a novel automated mammographic measure, V. Cancer Epidemiol Biomarkers Prev 2024; 33: 804-811.

[57] Huo CW, Hill P, Chew G, Neeson PJ, Halse H, Williams ED, Henderson MA, Thompson EW and Britt KL. High mammographic density in women is associated with protumor inflammation. Breast Cancer Res 2018; 20: 92.

[58] Esbona K, Inman D, Saha S, Jeffery J, Schedin P, Wilke L and Keely P. COX-2 modulates mammary tumor progression in response to collagen density. Breast Cancer Res 2016; 18: 35.

[59] Rice MS, Tamimi RM, Connolly JL, Collins LC, Shen D, Pollak MN, Rosner B, Hankinson SE and Tworoger SS. Insulin-like growth factor-1, insulin-like growth factor binding protein-3 and lobule type in the nurses' health study II. Breast Cancer Res 2012; 14: R44.

[60] Huh SJ, Oh H, Peterson MA, Almendro V, Hu R, Bowden M, Lis RL, Cotter MB, Loda M, Barry WT, Polyak K and Tamimi RM. The proliferative activity of mammary epithelial cells in normal tissue predicts breast cancer risk in premenopausal women. Cancer Res 2016; 76: 1926-1934.

[61] Tamimi RM, Colditz GA, Wang Y, Collins LC, Hu R, Rosner B, Irie HY, Connolly JL and Schnitt SJ. Expression of IGF1R in normal breast tissue and subsequent risk of breast cancer. Breast Cancer Res Treat 2011; 128: 243-250.

[62] Oh H, Eliassen AH, Wang M, Smith-Warner SA, Beck AH, Schnitt SJ, Collins LC, Connolly JL, Montaser-Kouhsari L, Polyak K and Tamimi RM. Expression of estrogen receptor, progesterone receptor, and Ki67 in normal breast tissue in relation to subsequent risk of breast cancer. NPJ Breast Cancer 2016; 2: 16032.

[63] Oh H, Eliassen AH, Beck AH, Rosner B, Schnitt SJ, Collins LC, Connolly JL, Montaser-Kouhsari L, Willett WC and Tamimi RM. Breast cancer risk factors in relation to estrogen receptor, progesterone receptor, insulin-like growth factor-1 receptor, and Ki67 expression in normal breast tissue. NPJ Breast Cancer 2017; 3: 39.

[64] Yaghjyan L, Heng YJ, Baker GM, Murthy D, Mahoney MB, Rosner B and Tamimi RM. Associations of stem cell markers in benign breast tissue with subsequent breast cancer risk. Am J Cancer Res 2023; 13: 6280-6289.

[65] Li R, Mukherjee MB, Jin Z, Liu H, Lin K, Liu Q, Dilger JP and Lin J. The potential effect of general anesthetics in cancer surgery: meta-analysis of postoperative metastasis and inflammatory cytokines. Cancers (Basel) 2023; 15: 2759.

[66] Liu H, Dilger JP and Lin J. Effects of local anesthetics on cancer cells. Pharmacol Ther 2020; 212: 107558.

[67] Liu H, Dilger JP and Lin J. Lidocaine suppresses viability and migration of human breast cancer cells: TRPM7 as a target for some breast cancer cell lines. Cancers (Basel) 2021; 13: 234.

[68] Li R, Xiao C, Liu H, Huang Y, Dilger JP and Lin J. Effects of local anesthetics on breast cancer cell viability and migration. BMC Cancer 2018; 18: 666.

Supplementary Table 1. Distribution of stem cell markers by BBD subtype (mean, SD, range)

Marker expression	Non-proliferative (n=195)	Proliferative without atypia (n=480)	Proliferative with atypia (n=60)
CD44 epithelium	45.7 (24.4) 0.47-95.9	49.5 (24.6) 0-97.2	47.1 (20.8) 3.1-95.7
CD44 stroma	8.2 (8.2) 0-43.0	8.7 (8.8) 0-50.6	8.1 (10.7) 0-49.9
CD24 epithelium	73.7 (16.5) 9.2-98.0	73.3 (17.0) 21.3-99.5	69.3 (18.6) 22.8-97.7
CD24 stroma	44.4 (18.2) 1.2-87.2	42.7 (19.3) 1.3-93.7	41.1 (18.6) 10.6-91.2
ALDH1A1 epithelium	14.0 (11.3) 0-61.0	14.5 (12.2) 0-59.2	15.7 (11.1) 1.1-43.9
ALDH1A1 stroma	23.4 (19.3) 0-81.1	26.5 (20.9) 0-93.8	31.0 (22.9) 0-83.6

Supplementary Table 2. Associations of reproductive factors with log-transformed expression of stem cell markers in benign breast biopsy samples (β coefficients and 95% Confidence intervals), adjusted for age and BMI only

Reproductive factor	CD44		CD24		ALDH1A1	
	In Epithelium	In Stroma	In Epithelium	In Stroma	In Epithelium	In Stroma
Nulliparity						
Nulliparous	0.16 (-0.03; 0.36)	0.02 (-0.37; 0.41)	-0.02 (-0.08; 0.04)	-0.09 (-0.21; 0.04)	-0.01 (-0.26; 0.23)	-0.17 (-0.60; 0.27)
Parous	Nulliparous	Reference	Reference	Reference	Reference	Reference
Breastfeeding, months						
0- <1	Reference	Reference	Reference	Reference	Reference	Reference
1- <12	0.02 (-0.20; 0.25)	-0.15 (-0.60; 0.29)	0.02 (-0.05; 0.09)	-0.11 (-0.25; 0.03)	0.13 (-0.14; 0.41)	0.15 (-0.32; 0.61)
12- <24	-0.5×10^{-2} (-0.24; 0.23)	-0.15 (-0.61; 0.31)	0.03 (-0.03; 0.11)	-0.06 (-0.20; 0.08)	0.25 (-0.04; 0.53)	0.31 (-0.17; 0.78)
≥24	-0.04 (-0.27; 0.20)	-0.21 (-0.68; 0.26)	0.04 (-0.05; 0.10)	-0.03 (-0.18; 0.11)	0.18 (-0.11; 0.47)	0.21 (-0.28; 0.69)
p-trend	0.64	0.48	0.52	0.75	0.26	0.41
Parity						
1	Reference	Reference	Reference	Reference	Reference	Reference
2	-0.23 (-0.44; 0.02)	-0.31 (-0.72; 0.17)	-0.01 (-0.08; 0.05)	-0.01 (-0.14; 0.11)	-0.07 (-0.32; 0.19)	-0.46 (-0.89; -0.04)
3	-0.18 (-0.41; 0.04)	-0.28 (-0.73; 0.17)	0.03 (-0.04; 0.10)	0.06 (-0.08; 0.20)	0.08 (-0.20; 0.36)	-0.04 (-0.50; 0.43)
≥4	-0.09 (-0.38; 0.20)	-0.05 (-0.64; 0.53)	0.06 (-0.03; 0.15)	0.14 (-0.14; 0.32)	-0.07 (-0.29; 0.43)	0.06 (-0.54; 0.66)
p-trend	0.75	0.66	0.08	0.24	0.79	0.42
Parity continuous	-0.03 (-0.11; 0.06)	-0.03 (-0.20; 0.14)	0.02 (-0.003; 0.05)	0.05 (-0.003; 0.10)	0.05 (-0.06; 0.15)	0.09 (-0.09; 0.26)
Age at 1st birth						
<25	Reference	Reference	Reference	Reference	Reference	Reference
25-29	-0.07 (-0.24; 0.10)	-0.20 (-0.55; 0.14)	-0.01 (-0.06; 0.05)	-0.11 (-0.21; -0.002)	-0.19 (-0.40; 0.02)	-0.36 (-0.71; 0.004)
≥30	0.06 (-0.14; 0.26)	-0.10 (-0.50; 0.30)	-0.02 (-0.09; 0.04)	-0.12 (-0.25; -0.001)	-0.27 (-0.51; -0.02)	-0.16 (-0.57; 0.25)
p-trend	0.59	0.58	0.44	0.04	0.03	0.40
Age at 1st birth, continuous (per 5 years)	0.02 (-0.06; 0.11)	-0.4×10^{-2} (-0.16; 0.17)	-0.02 (-0.04; 0.01)	-0.05 (-0.10; 0.001)	-0.10 (-0.20; 0.01)	-0.04 (-0.22; 0.13)
Age at menarche						
<12	0.07 (-0.12; 0.26)	-0.09 (-0.50; 0.33)	0.01 (-0.06; 0.08)	0.5×10^{-2} (-0.13; 0.13)	0.08 (-0.18; 0.34)	0.15 (-0.31; 0.62)
12	0.06 (-0.13; 0.25)	0.25 (-0.14; 0.64)	-0.02 (-0.08; 0.04)	-0.04 (-0.16; 0.08)	0.01 (-0.23; 0.26)	0.03×10^{-2} (-0.43; 0.43)
13	0.06 (-0.14; 0.27)	0.21 (-0.17; 0.60)	-0.01 (-0.07; 0.05)	0.02 (-0.10; 0.14)	-0.03 (-0.27; 0.21)	-0.06 (-0.49; 0.37)
>13	Reference	Reference	Reference	Reference	Reference	Reference
p-trend	0.64	0.73	0.93	0.75	0.46	0.47
Age at menarche, continuous (per 5 years)	-0.14 (-0.38; 0.11)	-0.4×10^{-2} (-0.48; 0.49)	-0.01 (-0.09; 0.07)	0.01 (-0.14; 0.16)	-0.12 (-0.42; 0.18)	-0.23 (-0.77; 0.32)
Time between menarche and age at 1st birth, continuous (per 5 years)	0.03 (-0.05; 0.11)	0.01 (-0.15; 0.16)	-0.01 (-0.04; 0.01)	-0.04 (-0.09; 0.01)	-0.07 (-0.18; 0.02)	-0.02 (-0.18; 0.14)
Time since last birth, continuous (per 5 years)	-0.02 (-0.10; 0.07)	0.03 (-0.14; 0.20)	0.05×10^{-2} (-0.03; 0.03)	0.02×10^{-2} (-0.05; 0.05)	0.02 (-0.09; 0.12)	-0.07 (-0.24; 0.11)