

Increased promoter methylation in exfoliated breast epithelial cells in women with a previous breast biopsy

Eva P. Browne,¹ Elizabeth C. Punska,¹ Sarah Lenington,¹ Christopher N. Otis,² Douglas L. Anderton³ and Kathleen F. Arcaro^{1,*}

¹Department of Veterinary and Animal Science; ²Social and Demographic Research Institute; University of Massachusetts at Amherst; Amherst, MA USA;

³Department of Pathology; Baystate Medical Center; Springfield, MA USA

Key words: promoter methylation, pyrosequencing, breast milk, biopsy, *GSTP1*, *RASSF1*, *SFRP1*, breast epithelial cells, breast cancer risk

Abbreviations: CpG, cytosine phosphodiester bonded to guanine; *GSTP1*, glutathione-S-transferase pi; MACS, magnetic antibody cell separation; *RASSF1*, rass association domain family 1; *SFRP1*, secreted frizzle related protein 1

Accurately identifying women at increased risk of developing breast cancer will provide greater opportunity for early detection and prevention. DNA promoter methylation is a promising biomarker for assessing breast cancer risk. Breast milk contains large numbers of exfoliated epithelial cells that are ideal for methylation analyses. Exfoliated epithelial cells were isolated from the milk obtained from each breast of 134 women with a history of a non-proliferative benign breast biopsy (Biopsy Group). Promoter methylation of three tumor suppressor genes, *RASSF1*, *SFRP1* and *GSTP1*, was assessed by pyrosequencing of bisulfite-modified DNA. Methylation scores from the milk of the 134 women in the Biopsy Group were compared to scores from 102 women for whom a breast biopsy was not a recruitment requirement (Reference Group). Mean methylation scores for *RASSF1* and *GSTP1* were significantly higher in the Biopsy than in the Reference Group. For all three genes the percentage of outlier scores was greater in the Biopsy than in the Reference Group but reached statistical significance only for *GSTP1*. A comparison between the biopsied and non-biopsied breasts of the Biopsy Group revealed higher mean methylation and a greater number of outlier scores in the biopsied breast for both *SFRP1* and *RASSF1*, but not for *GSTP1*. This is the first evidence of CpG island methylation in tumor suppressor genes of women who may be at increased risk of developing breast cancer based on having had a prior breast biopsy.

Introduction

It is estimated 230,480 women will be diagnosed with breast cancer and 39,520 will die from the disease in the US in 2011.¹ Despite significant advances in early detection and improved treatment, breast cancer remains the second most diagnosed cancer and the second leading cause of cancer deaths in women.¹ Recent guidelines by the US Preventive Services Task Force recommend screening mammography begin at age 50 rather than the previously recommended age 40 for all women at average risk and that “the decision to start regular, biennial screening mammography before the age of 50 years should be an individual one and take patient context into account, including the patient’s values regarding specific benefits and harms.”² This recommendation assumes the individual has an understanding of her risk of developing breast cancer.

Epidemiological studies have consistently identified established risk factors, including early menarche, nulliparity, late age at first birth, previous biopsy and family history as contributing

to breast cancer incidence. However, these same studies find that almost all women have at least one of the established risk factors.³⁻⁶ Furthermore, while women with a family history of breast cancer are at increased risk, most women with a family history of breast cancer do not develop the disease. A collaborative reanalysis of 52 epidemiological studies found that 8 out of 9 women who develop breast cancer do not have an affected mother, sister or daughter.⁷ The NCI estimates that only 5–15% of all breast cancers are associated with mutations in autosomal dominant genes such as *BRCA1*, *BRCA2*, *Tp53* and *PTEN*.^{1,7} The vast majority of breast cancers are sporadic, making it extremely difficult for women to assess their individual breast cancer risk.

There is a great desire and need to find molecular markers that accurately predict individual breast cancer risk. DNA methylation of tumor suppressor genes is considered one of the most promising of the molecular biomarker candidates.^{8,9} Methylation of CpG islands in promoter regions frequently results in transcriptional silencing of the gene and is thought to occur early in the etiology of breast cancer.¹⁰ Most importantly, CpG island

*Correspondence to: Kathleen Arcaro; Email: karcaro@vasci.umass.edu

Submitted: 08/18/11; Revised: 09/29/11; Accepted: 10/02/11

DOI: 10.4161/epi.6.12.18280

methylation is potentially reversible with drug treatment or diet modification.¹¹ However, the promoter methylation pattern characteristic of cancer, early stage disease or increased risk is tissue specific. That is to say, methylation associated with breast cancer risk must be assessed in breast tissue.

Studies examining the relationship between promoter methylation and breast cancer risk have used DNA from cells in nipple aspirate, ductal lavage and random periareolar fine-needle aspirate (PFNA).¹²⁻¹⁵ While these studies have provided important information, there are several drawbacks to these sampling methods such as, the amount of material obtained being small (e.g., nipple aspirate and ductal lavage), the material obtained being from a limited area or single lobule of the breast (all sampling methods) and method of collection being invasive and not likely to be widely accepted (PFNA and ductal lavage).^{14,15} In contrast, breast milk contains millions of epithelial cells representing all of the tissue in the entire mammary gland and can be easily obtained from each breast as exfoliated epithelial cells are naturally present in breast milk. Women are likely to participate in studies of breast milk because collecting milk is non-invasive. Additionally, women who have multiple pregnancies can contribute samples over a number of years, providing important information regarding methylation changes over time.

Promoter methylation of the three genes analyzed in the present study has been detected previously in breast cancer, and may increase a woman's risk of developing this disease. *RASSF1* (*Rass association domain family1 protein*), a tumor suppressor gene implicated in the development of breast cancer,¹⁶ promotes apoptosis and cell cycle arrest, and reduces the tumorigenicity of cancer cell lines.¹⁷ Aberrant promoter methylation of *RASSF1* is frequently detected in breast cancer tissue¹⁸⁻²¹ and has also been detected in serum of breast cancer patients²² and in fine needle aspirate from benign epithelium of women at high risk for breast cancer.²³ *SFRP1* (*Secreted frizzles related protein 1*) is a tumor suppressor gene encoding a WNT signaling antagonist abundantly expressed in normal breast tissue. *SFRP1* has been reported to be hypermethylated in >70% of breast cancers.²⁴ *GSTP1* (*Glutathione-S-transferase P1*) plays a role in protecting cells from cytotoxic and carcinogenic agents.²⁵ Expression of *GSTP1* varies from tissue to tissue; however, loss of *GSTP1* expression has been attributed to DNA hypermethylation.²⁶ Methylation of *GSTP1* has been detected in breast cancer tissue^{18,27-30} as well as breast fluids from cancer patients^{3,31} and has been correlated with age in benign prostate.³²

We have chosen to study methylation changes in candidate biomarker genes (*RASSF1*, *SFRP1* and *GSTP1*) and their relationship to breast cancer risk using milk from nursing women. Since most women do not develop breast cancer, the vast majority of nursing mothers are not expected to have increased CpG island promoter methylation. Accordingly, we previously demonstrated that mean methylation levels were extremely low in the exfoliated epithelial cells isolated from breast milk of women at no known increased risk of developing breast cancer.³³ To determine whether methylation analysis of cells in breast milk can be used to assess breast cancer risk it is important to select a population in which it is probable that a subset of participants will

be at a significantly increased risk for developing breast cancer. Our approach is to recruit nursing mothers who either have had a breast biopsy or are scheduled to have a breast biopsy due to a suspicious lump. This is an ongoing study in which we continue to recruit women and acquire annual follow-up information. In the present report, we compare the methylation scores obtained from cells in the milk of 134 women whose breast biopsy indicated the lowest risk, benign, non-proliferative lesion, and compare these methylation scores with those from a reference group who were not selected based on biopsy history. Among women who had a breast biopsy we also compare the methylation scores obtained from their biopsied and non-biopsied breasts. Results show increased methylation in the Biopsy Group and considerable individual variability. Only long-term follow-up will reveal if these individual methylation scores are accurate predictors of breast cancer risk.

Results

Subject recruitment. In the Biopsy Group, milk samples were collected and processed from 141 consented women whose biopsy reports indicated they had a non-proliferative lesion (Category 1). The majority of biopsies were surgical and core (47 and 38%, respectively) followed by FNA (10%), and by far the most frequent diagnosis was fibroadenoma (40%) followed by lactational changes, fibrocystic change and fibrosis (Table 1). Of the ten women who received a "negative" diagnosis, five of these diagnoses were based on a FNA biopsy and were considered insufficient to retain the women in the low risk Category 1. Of the remaining 135 women in Category 1, epithelial-enriched cell fractions were obtained from the milk of at least one breast from 134 women. For the Reference Group, milk samples were obtained from 111 consented women and epithelial-enriched cell fractions were obtained from the breast milk (combined from both breasts) of 102 women (see Wong et al.³³). Therefore, all following analyses are based on 134 women for the Biopsy Group and 102 women for the Reference Group.

Of the women who provided milk for the Biopsy Group, the majority (72%) was recruited through the Army of Women (www.armyofwomen.org). The remainder was recruited through various websites and by sending brochures to mammography centers and lactation consultants. The 134 women in the Biopsy Group resided in 36 states including CA (12), CO (2), CT (3), FL (4), GA (5), HI (1), IA (4), IL (10), IN (1), KS (2), KY (1), MA (16), MD (6), ME (1), MI (5), MN (3), MO (1), MT (1), NC (4), ND (1), NE (1), NJ (5), NM (2), NY (8), OH (1), OK (1), OR (1), PA (2), RI (1), SC (2), TN (1), TX (9), VT (1), VA (6), WA (6) and WI (4).

Of the women who provided milk for the Reference Group, the majority (~80%) lived within 20 miles of Amherst, MA and were recruited through local advertisements and fliers. The remaining roughly 20% of women lived within 100 miles of Amherst, MA or were visiting Amherst at the time they donated a milk sample.

Subject demographics. Neither mean age, age at first birth, baby's age at time of milk donation, number of live births, nor

BMI differed significantly between the 134 women in the Biopsy Group and the 102 women in the Reference Group (Table 2). The majority of women in both the Biopsy and Reference Groups were Caucasian (88 and 90%, respectively). There was a small percentage of Hispanics (7 and 1%, respectively) followed by Asian and Pacific Islanders (2 and 6%, respectively) and African Americans (2 and 1%, respectively) in each of the groups. The only demographic that differed significantly between the Biopsy and Reference Groups was family history of breast cancer. Thirty percent of the women in the Biopsy Group had a first degree female relative (mother or sister) with breast cancer compared to only 7% in the Reference Group (Fisher's Exact $p = 0.000$), and 63% of women in the Biopsy Group had any family history of breast cancer compared to 45% in the Reference Group (Fisher's Exact $p = 0.006$).

Milk sample demographics. The major difference in the breast milk samples between the two study groups was the origin of the milk. In the Reference group, the single milk sample from each woman came from either one breast or both breasts, but this information was not obtained from each woman. In contrast, every milk sample in the Biopsy Group came from one breast only, the origin of the milk (left or right breast) was recorded and all but one woman provided milk samples from both breasts. The mean volumes of the milk samples were similar between the biopsied and non-biopsied breasts from all women in Biopsy Group (59 and 57 mL, respectively) and yielded similar numbers of epithelial-enriched cells, and ng of DNA (Table 3). In contrast, the volume of the combined milk samples in the Reference Group were 48% greater (mean = 86 mL) but yielded roughly 100% more epithelial cells and 290% more DNA (Table 3). Note that there are differences between Group sample size and demographic sample size in Table 3. In several cases the volume of milk was not recorded, the cells in the epithelial-enriched fraction were not counted, and the DNA was not quantified.

To determine if having a recent biopsy altered the sample demographics of the Biopsy Group, we compared the volume of milk, number of epithelial cells and ng of DNA obtained from 30 women who had a biopsy within one year of milk donation to those who had a biopsy greater than one year apart from their milk donation (Table 3). There were no significant differences in any of the sample parameters regardless of the time since biopsy or whether the sample was from the biopsied or non-biopsied breast.

Comparison of methods for assessing methylation in the biopsy and reference groups. For both the Biopsy and Reference Groups, the cell separation, DNA extraction, bisulfite treatment and PCR amplification were conducted under similar conditions in our laboratory as described in the methods. The pyrosequencing, however, was conducted in two different laboratories: for the Biopsy Group the sequencing was conducted in our laboratory on a PyroMark Q24, while the pyrosequencing for the Reference Group was conducted in a commercial laboratory on a PSQ96 HS. To address the potential impact of differences in pyrosequencing equipment/methods for Biopsy and Reference Groups, pyrosequencing was repeated for selected samples from the Reference Group using the same equipment

Table 1. Summary of biopsy results for the 134 women in Category 1

Biopsy types; n (%)		
FNA	14	(10.45)
Core	51	(38.06)
Surgical	63	(47.01)
Unknown	6	(0.04)
Biopsy diagnoses; n (%)		
Adenoma	1	(0.75)
Adenosis	3	(2.24)
Adipose tissue	2	(1.50)
Angiolipoma	1	(0.75)
Cyst	8	(5.97)
Dense stroma	1	(0.75)
Diffuse cystic mastopa	1	(0.75)
Duct ectasia	1	(0.75)
Fat Necrosis	1	(0.75)
Fibrocystic change	13	(9.70)
Fibroadenoma	54	(40.30)
Fibrosis	12	(8.96)
Galactocele	2	(1.50)
Granuloma	1	(0.75)
Haramtoma	1	(0.75)
Hemangioma	1	(0.75)
Inflammatory cyst	1	(0.75)
Inflammatory lymph node	2	(1.50)
Lactating adenoma	2	(1.50)
Lactational changes	14	(10.45)
Lipoma	1	(0.75)
Mastitis	5	(3.73)
Negative	5*	(3.73)
Phyllodes tumor	1	(0.75)

*An additional five women received a negative diagnosis based on FNA, which was considered insufficient to retain them in the low risk Category 1.

used to perform pyrosequencing for the Biopsy Group. In 2011, archived bisulfite-modified DNA from 22 epithelial-enriched cell samples were retrieved from the minus 80°C freezer amplified for *RASSF1*, *SFRP1* and *GSTP1* and products were sequenced on the PyroMark Q24. Figure 1 shows the similarity of CpG-specific and overall mean methylation scores of two representative samples sequenced on the PyroMark in 2011 and the PSQ96 in 2008. For the 22 samples, the overall mean methylation scores for *RASSF1*, *SFRP1* and *GSTP1* did not differ between samples sequenced in 2011 and the same samples sequenced in 2008 (*RASSF1*: 4.75 (2011) vs. 4.45 (2008); $t = 0.64$; $df = 21$; $p = 0.53$; *SFRP1*: 9.21 (2011) vs. 6.53 (2008); $t = 1.41$; $df = 19$; $p = 0.09$, and *GSTP1*: 3.13 (2011) vs. 3.72 (2008); $t = -1.33$; $df = 11$; $p = 0.21$). Furthermore, an ANOVA revealed no CpG site-specific differences for *RASSF1* between the samples analyzed in 2008 and again in 2011 ($F = 0.96$; $df = 8, 1$; $p = 0.33$). However, there

Table 2. Percent is based on the number of women from each group who responded

Subject demographics	Biopsy group		Reference group	
Participants; n	134		102	
Age at milk donation; mean (S.D.) and range in years	34.0 (4.6)	23–52	32.3 (5.6)	19–45
Age at biopsy; mean (S.D.) and range in years	29.7 (6.0)	13–51	NA	NA
Years since biopsy; mean (S.D.) and range in years	4.3 (3.9)	-0.3-15.9	NA	NA
Age at first birth; mean (S.D.) and range in years	30.6 (4.6)	18–52	29.8 (5.2)	19–43
Baby's age at milk donation; mean (S.D.) and range in days	277.3 (230.4)	30–1440	246.3 (188.7)	30–900
Number of live births; mean (S.D.) and range	1.9 (0.88)	1–5	1.6 (0.83)	1–5
Body mass index; mean (range)	23.8 (4.1)	18.2–39.3	24.8 (5.6)	16.2–59.3
Race/Ethnicity; n (%)				
Caucasian	118 (88.06)		90 (90)	
Hispanic	9 (6.72)		1 (1)	
Asian and Pacific Islander	3 (2.24)		6 (6)	
African American	3 (2.24)		1 (1)	
Other	1 (.75)		2 (2)	
First degree female relative with breast cancer; n (%)	39 (29.1)		7 (6.7)	
Any family history of breast cancer; n (%)	85 (63.4)		46 (45.1)	

There were missing data from 1–5 women for either age at first birth, baby's age at milk donation, BMI or race/ethnicity; NA, not applicable.

Table 3. Sample demographics

	Reference group (n = 102)		Biopsy group (n = 134)				Women in biopsy group who had a biopsy within one year of donating a milk sample			
			Biopsied breast		Non-biopsied breast		Biopsied breast		Non-biopsied breast	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
Milk (volume in mL)	102	86 (33)	126	59 (26)	131	57 (28)	28	56 (31)	30	56 (32)
Cell number in epithelial-enriched fraction (x10 ⁶)	101	2.5 (3.0)	126	1.2 (3.1)	130	1.4 (5.6)	28	0.9 (2.5)	30	0.7 (1.0)
DNA (ng)	94	2591 (2919)	117	668 (517)	122	661 (622)	27	626 (431)	29	684 (570)

See text for an explanation of differences in sample size.

was a significant difference for *SFRPI* between the samples analyzed in 2008 and again in 2011 ($F = 5.04$; $df = 7, 1$; $p = 0.025$) and for *GSTPI* ($F = 8.60$; $df = 12, 1$; $p = 0.004$). It is possible that the CpG site-specific differences between the two analyses were related to the length of time the bisulfite-modified DNA was archived (over two years) or the number of times the aliquot was thawed. Regardless of the reason for the site-specific differences, we limited all comparisons between the two study groups to the overall mean methylation scores, as these values did not differ between the same samples analyzed in 2008 on the PSQ96 HS and in 2011 on the PyroMark Q24.

Differences in CpG promoter methylation in epithelial cells from the breast milk of women in the biopsy versus the reference group. We examined the methylation levels of 30 individual CpG sites within promoter regions of three tumor suppressor genes. As can be seen in **Figure 2**, the two populations differ in methylation scores. The most dramatic difference between the two populations is the greater number of outlier scores in the Biopsy Group, particularly for *GSTPI*. This might be expected given that the sample sizes in the Biopsy Group (individual

breasts) were roughly twice that of the Reference Group (single sample per woman). Therefore we examined the percentage of population outliers by calculating the scores that were greater than the 75th percentile + 1.5*IQR of the combined Biopsy and Reference Groups. The percentage of outlier scores is significantly greater in the Biopsy Group than in the Reference Group for *GSTPI* (16.3 vs. 0%; Fisher's Exact $p < 0.00$) but not for either *RASSFI* (9.7 vs. 5.9%) or *SFRPI* (7.8 vs. 5.9%). While less dramatic than the outliers, the overall means also are significantly higher in the Biopsy Group than in the Reference Group for both *RASSFI* (7.00 vs. 4.72; $t = 2.93$; $p = 0.002$) and *GSTPI* (9.06 vs. 3.64; $t = 5.14$; $p = 0.00$), but not for *SFRPI* (6.29 vs. 5.80; $t = 0.63$; $p = 0.27$).

Next, to determine the extent to which the differences between the two study groups were related to demographic covariates, we conducted pooled OLS-ANOVA comparisons including study group, age, age at first birth, number of live births, baby's age (as a surrogate for length of nursing), BMI and family history of breast cancer (first degree female relative with breast cancer and any family history of breast cancer were considered separately).

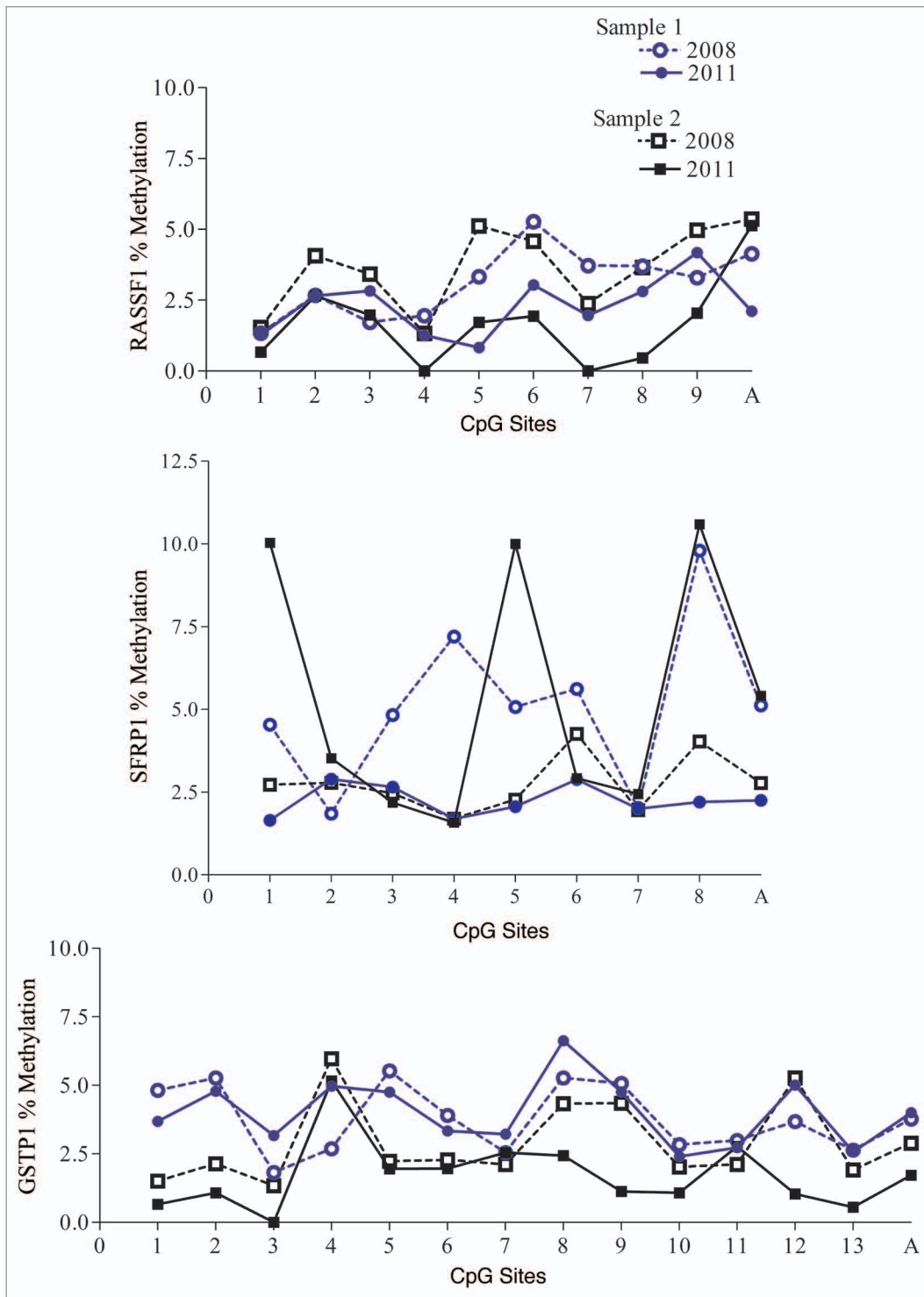


Figure 1. Comparison of the pyrosequencing methods used for methylation analyses in the Biopsy and Reference Groups. Gene-specific PCR amplification and pyrosequencing was performed on a total of 22 archived samples of bisulfite-modified DNA stored for over two years at minus 80°C. Results from two representative samples of bisulfite-modified DNA sequenced in 2008 on the PSQ96 HS and again in 2011 on the PyroMark Q24 are shown for *RASSF1* (top part), *SFRP1* (middle part) and *GSTP1* (bottom part). Filled symbols and solid lines indicate samples sequenced in 2011; open symbols and dashed lines indicate samples sequenced in 2008; A = average of all CpG sites per gene.

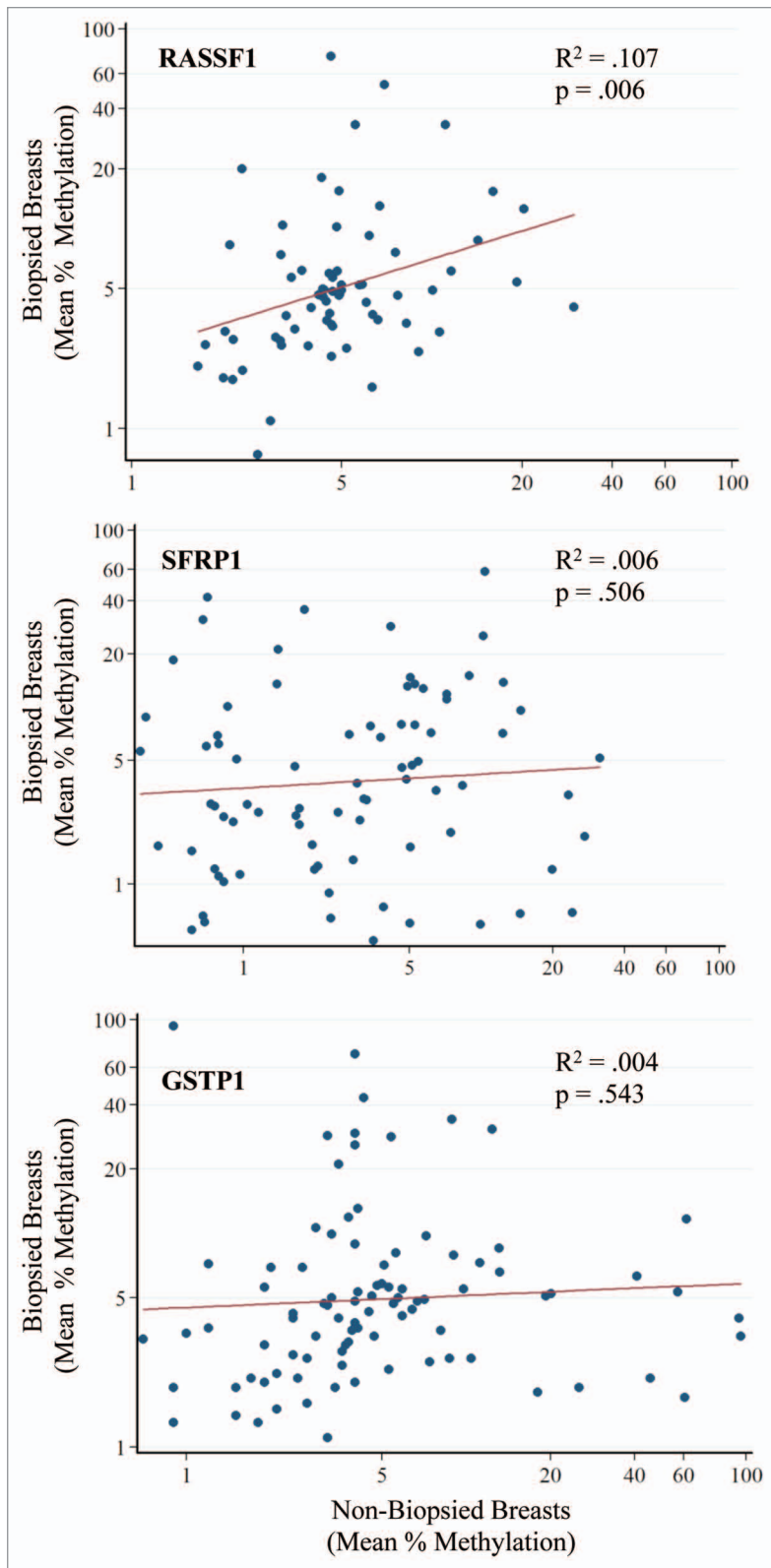


Figure 3. Comparison of biopsied and non-biopsied breasts. The correlation between mean CpG methylation scores of the biopsied and non-biopsied breasts is presented for women who had a biopsy in only one breast and had some methylation data from each breast. Log of mean methylation scores is shown for *RASSF1* (top), *SFRP1* (middle) and *GSTP1* (bottom).

(1) the group of women with a non-proliferative breast biopsy diagnosis had a slightly increased probability of CpG promoter methylation in two of three genes examined as compared to a group for whom a breast biopsy was not a requirement; (2) the methylation patterns differed between the biopsied and non-biopsied breasts for two of the three genes and (3) while the mean methylation scores were low for most women, a subset of women had significantly higher CpG promoter methylation.

The difference in mean methylation scores and percentage of outliers between the Biopsy and Reference Groups is strong evidence that the methylation signal is providing important information about the breast and is not simply indicating the breast is lactating. Both groups of epithelial cells come from the breasts of lactating women of similar ages and reproductive history. The major difference between the groups is the recruitment strategy and selection criteria of a previous biopsy. The women in the Biopsy Group are a self-selected group of nursing mothers residing in 36 different states, the majority of whom responded to Internet recruiting strategies describing a study on molecular markers associated with breast cancer risk. As a study population they are unusual in that 30% report having a mother or sister with breast cancer. This percentage is much higher than the population average of 5–7%,^{34,35} suggesting that these women may have been highly motivated to participate in a breast cancer study because of a family history of breast cancer.

While a greater percentage of women in the Biopsy Group had a first-degree female relative with breast cancer, family history was not significantly associated with increased methylation. Indeed, the absence of family history of breast cancer was associated with a significant increase in promoter methylation for *GSTP1*. The lack of correlation between family history and increased methylation is consistent with studies reporting that most breast cancers occur in women without a family history of breast cancer.⁷ That little of the variability in methylation scores is explained by established risk factors is expected given the low discriminatory accuracy of most breast cancer risk models.

In normal cells, *RASSF1*, *SFRP1* and *GSTP1* are involved in controlling cell cycle, repairing DNA and metabolizing xenobiotics.^{17,25} They were selected for methylation analysis in the present study because they have been shown to be transcriptionally silenced by promoter methylation and are frequently methylated in breast cancer tissue.^{18–20,22–24} It is important to keep in mind that the majority of women in the Biopsy Group have low methylation scores. While the overall mean methylation scores are above 5%, greater than 50% of the women are below 5% (overall mean methylation

