

DNA promoter hypermethylation in nipple fluid: a potential tool for early breast cancer detection

Jolien S. de Groot¹, Cathy B. Moelans¹, Sjoerd G. Elias², Mary Jo Fackler³, Robert van Domselaar¹, Karijn P.M. Suijkerbuijk¹, Arjen J. Witkamp⁴, Saraswati Sukumar³, Paul J. van Diest¹, Elskan van der Wall⁵

¹Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

²Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

³Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, USA

⁴Department of Surgery, University Medical Center Utrecht, Utrecht, The Netherlands

⁵Department of Medical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

Correspondence to: Elskan van der Wall, **e-mail:** e.vanderwall@umcutrecht.nl

Keywords: nipple fluid, methylation, breast cancer

Received: December 10, 2015

Accepted: February 29, 2016

Published: March 25, 2016

ABSTRACT

Introduction: Nipple fluid aspiration provides direct non-invasive sampling of fluid from the mammary ductal system, where the majority of breast cancers originate. DNA promoter hypermethylation (“methylation”) occurs early and at high frequency in breast carcinogenesis, bearing the potential as a biomarker for cancer detection at its earliest stages. We assessed methylation in nipple fluid from breasts of healthy women, of women with sporadic breast cancer and of their contralateral breasts. Our goal was to investigate whether nipple fluid can be used as a reliable methylation biomarker source.

Methods: Methylation levels of 13 genes were analysed by quantitative multiplex-methylation specific PCR (QM-MSP) in nipple fluid samples from breasts of healthy women, and from the affected and contralateral breasts of breast cancer patients.

Results: Methylation analysis of the low-volume nipple fluid samples was feasible. Despite the generally low methylation levels, cancerous and healthy breasts nipple fluid could be discriminated with an area under the receiver operating characteristic curve (AUC) of 0.64 ($p < 0.01$) based on a multivariate model including *AKR1B1*, *ALX1*, *RASSF1A* and *TM6SF1*. Within-patient differences between cancerous and contralateral nipple fluid samples were less prominent.

Conclusions: Cancerous nipple fluid contains increased levels of methylation of tumor suppressor genes that potentially could serve as a biomarker for early breast cancer detection.

INTRODUCTION

Worldwide, breast cancer is the most common cancer in women [1, 2]. Five-year survival rates range from more than 80% in developed countries to less than 40% in developing areas [1]. Early detection improves breast cancer survival, resulting in the implementation of various imaging modalities for screening, mammography being the most commonly applied modality.

An intraductal approach might evolve into an alternative screening method by offering a way to directly

sample fluid from the mammary ductal system, where breast cancer originates in the majority of patients. Aspirated fluid contains cells, DNA, RNA and proteins directly derived from the breast ducts and can thereby be a rich source of breast cancer biomarkers [3, 4]. Fluid from the breast can be obtained by invasive techniques like random fine needle aspiration (FNA) or ductal lavage (DL), but nipple fluid can also be obtained in a completely non-invasive way by an oxytocin-assisted nipple fluid aspiration under vacuum (NAF). Besides being less invasive, NAF causes less discomfort and is easier to perform compared to invasive

techniques [5]. We have previously shown that, with this technique, nipple fluid can be obtained successfully and without discomfort in healthy women and women at increased risk of breast cancer [6, 7].

To improve breast cancer screening, the detection of DNA promoter hypermethylation (further denoted “methylation”) in nipple fluid could be of added value. Methylation of tumor suppressor genes occurs early and at high frequency in most cancer types [7]. The detection of methylation in ductal fluids is feasible using a very sensitive PCR technique like quantitative multiplex methylation-specific PCR (QM-MSP) [8] and is therefore a promising biomarker for breast cancer screening [4]. Previously, we described that methylation of 11 genes (*RARB*, *RASSF1*, *TWIST1*, *CCND2*, *ESSR1*, *SCGB3A1*, *BRCA1*, *BRCA2*, *CDKN2A*, *APC* and *CDHI*) is involved in carcinogenesis of sporadic and *BRCA1*-associated breast cancer [9]. Subsequently we validated a new set of genes in sporadic breast cancer tissue and demonstrated that the promoters of the *AKR1B1*, *ALX1*, *GHSR*, *GREM1*, *RASGRF2*, *SFRP2*, *TM6SF1* and *TMEFF2* genes were significantly differentially methylated in normal versus malignant breast tissues [10-12]. As a first step to test the screening potential of methylation in nipple fluid, we compared the methylation status of a subset of these genes in nipple fluid samples obtained from breasts of healthy women, and from the affected and contralateral breasts of breast cancer patients.

RESULTS

Nipple fluid aspiration and baseline characteristics

We used 88 nipple fluid samples from 49 healthy women to set methylation thresholds. In total 100 breast cancer patients were included in the study. In 3 patients the breast cancer was bilateral and one patient was excluded because of missing data on if the yielding breast was affected. Thus, out of 103 attempts to collect nipple fluid from affected breasts and 97 from contralateral breasts, we were able to collect fluid from 55 (53.4%) and 66 (68.0%), respectively. Methylation analysis could be performed in 54 samples from the affected breast and in 39 contralateral samples. After analysis of the surgical specimen, no *in situ* or invasive breast cancer was detected in two women and these samples were consequently excluded from further analysis. Definitive analysis could therefore be performed on 52 breast cancer and 39 contralateral nipple fluid samples, of which 31 were paired. The flow chart of this process is shown in Figure 1.

Baseline characteristics of all included women are shown in Table 1. Mean age of healthy women was 30.7 years (standard deviation (SD) 12.0, median 25.0). In breast cancer patients from whom nipple fluid was collected at the affected or contralateral side, the mean

age was 54.6 years (SD 10.4, median 54.0) and 54.9 years (SD 10.0, median 56.0), respectively. The age of affected women significantly differed from the age in the healthy control group ($p < 0.001$ for both). Other clinical parameters correlating with age were also significantly different between these categories, such as history of breastfeeding, parity, oral contraceptive use and menopausal status ($p < 0.001$ for all, in samples with cancer as well as contralateral samples). Moreover, women with breast cancer significantly more often suffered from spontaneous nipple discharge ($p = 0.024$). Pathological characteristics of the breast tumors in the fluid yielding breasts are shown in Supplementary Table S1. In the breast cancer group, one woman carried a *BRCA1* mutation and one woman a *BRCA2* mutation.

Nipple fluid obtained from the affected or contralateral breasts from breast cancer patients was more often green (23% and 16% versus 2%) or brown/red/bloody (6% and 5% versus 2%) than nipple fluid from healthy women ($p < 0.001$ and $p = 0.019$, respectively). The volume of the aspirated nipple fluid in healthy controls was in 66.0% of the samples up to 10 μ l, in 23.4% 10-50 μ l, and in 10.6% more than 50 μ l. For the affected breasts, these numbers were 72.5%, 23.5% and 4.0%, and for contralateral breasts 59.0%, 30.7% and 10.3%, respectively.

Methylation in nipple fluid of healthy volunteers versus breast cancer patients

We used the 90th methylation percentile in the healthy samples as a cut-off for defining the presence of methylation in the nipple fluid samples. This cut-off is shown for each gene in the 13 gene panel in Table 2A, together with the number of missing values, the number of methylated samples in the 52 breast cancer and 49 healthy control samples, the univariable odds ratio (OR) and the corresponding p-value per gene. We included the 4 genes with lowest amount of missing data in the bilateral sample set, i.e. *AKR1B1*, *ALX1*, *RASSF1A* and *TM6SF1*, in multivariable analysis (Table 2B). This panel predicted the presence of breast cancer with an AUC of 0.64 (95%CI 0.54 - 0.74, $p < 0.01$). The graphs of the calibration and ROC curve corresponding to this model are shown in Figure 2A and 2B. At 90% specificity, sensitivity of the 4 gene panel was 31%. Using non-missing (non-imputed) data only, the 4-gene panel was associated with breast cancer occurrence with an AUC of 0.63 (95%CI 0.50 - 0.76, $p = 0.043$). For the CMI, the AUC was 0.68 (95%CI 0.56 - 0.81, $p < 0.01$), see also Figure 2C.

Methylation in nipple fluid of breasts with cancer versus contralateral breasts

Baseline characteristics of the 31 women included in the paired analysis were comparable with the

non-selected group of breast cancer patients (data not shown). Mean age of these women was 54.5 years (95%CI 50.4 – 58.5, median 54.0). Table 3A shows the results of univariable analysis for the 13 separate genes in the breast cancer samples versus their contralateral samples. The gene panel consisting of *AKR1B1*, *ALX1*, *RASSF1A* and *TM6SF1* was used in multivariable

analysis, of which the results are shown in Table 3B. This panel predicted the presence of breast cancer with an AUC of 0.58 (95%CI 0.44 - 0.72, p=0.26). The corresponding calibration and ROC curves are shown in Figure 2D and 2E. At 90% specificity, sensitivity was 21%. Using non-missing (non-imputed) data only, the 4-gene panel predicted breast cancer occurrence with

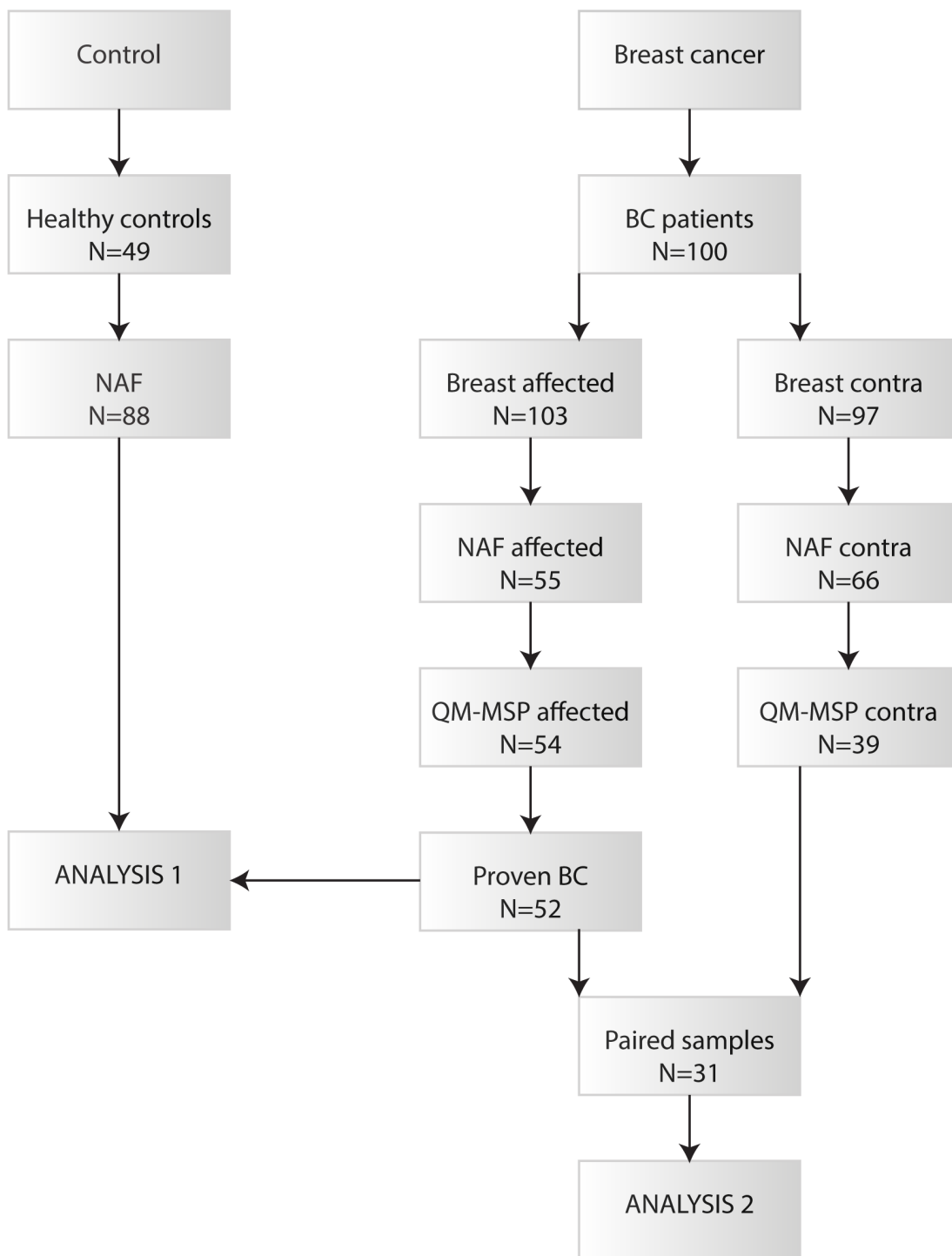


Figure 1: Flow chart of the collected nipple fluid samples that have been analysed by QM-MSP.

Table 1: Baseline characteristics of healthy volunteers and breast cancer patients

Feature		Nipple fluid healthy volunteers N=49 (valid %)	Nipple fluid breast cancer N=52 (valid %)	Nipple fluid contralateral breast N=39 (valid %)
Breast size	A-B	13 (28.9%)	15 (28.8%)	11 (28.2%)
	C-D	26 (57.8%)	31 (59.6%)	21 (53.8%)
	>D	6 (13.3%)	6 (11.5%)	7 (17.9%)
	Not available	4	0	0
Spontaneous nipple discharge	No	41 (87.2%)	38 (73.1%)	28 (71.8%)
	Yes	6 (12.8%)	14 (26.9%)	11 (28.2%)
	Not available	2	0	0
Breast feeding in history	No	37 (78.7%)	23 (44.2%)	17 (43.6%)
	Yes	10 (21.3%)	29 (55.8%)	22 (56.4%)
	Not available	2	0	0
Duration breast feeding	< 6 months	2 (4.3%)	12 (23.1%)	9 (23.1%)
	6 – 12 months	7 (14.9%)	7 (13.5%)	6 (15.4%)
	>12 months	1 (2.1%)	10 (19.2%)	7 (17.9%)
	Not applicable	37 (78.7%)	23 (44.2%)	17 (43.6%)
	Not available	2	0	0
Parity	Nulliparous	33 (70.2%)	14 (26.9%)	7 (17.9%)
	Parous	14 (29.8%)	38 (73.1%)	32 (82.1%)
	Not available	2	0	0
Age at first birth	<20 years	0 (0.0%)	4 (7.8%)	4 (10.5%)
	20 – 24 years	3 (6.4%)	8 (15.7%)	6 (15.8%)
	25-29 years	2 (4.3%)	10 (19.6%)	11 (28.9%)
	>30 years	9 (19.1%)	16 (31.4%)	11 (28.9%)
	Not applicable	33 (70.2%)	13 (25.5%)	6 (15.8%)
	Not available	2	1	1
Current use oral contraceptive	No	22 (46.8%)	52 (100.0%)	39 (100.0%)
	Yes	25 (53.2%)	0 (0.0%)	0 (0.0%)
	Not available	2	0	0
Current hormonal replacement therapy	No	47 (100.0%)	51 (98.1%)	39 (100.0%)
	Yes	0 (0.0%)	1 (1.9%)	0 (0.0%)
	Not available	2	0	0
Age at menarche	≤ 13 years	25 (54.3%)	34 (66.7%)	24 (63.2%)
	>13 years	21 (45.7%)	17 (33.3%)	14 (36.8%)
	Not available	3	1	1
Menopausal status	Premenopausal	41 (87.2%)	25 (48.1%)	15 (38.5%)
	Postmenopausal	6 (12.8%)	27 (51.9%)	24 (61.5%)
	Not available	2	0	0
Menstrual cycle day	Day 1 – 7	13 (27.7%)	8 (16.0%)	5 (13.2%)
	Day 7 – 14	5 (10.6%)	4 (8.0%)	2 (5.3%)
	Day 14 – 21	14 (29.8%)	1 (2.0%)	0 (0.0%)
	Day 21 – 28	5 (10.6%)	5 (10.0%)	3 (7.9%)
	Day >28	4 (8.5%)	4 (8.0%)	3 (7.9%)
	IUD	0 (0.0%)	1 (2.0%)	1 (2.6%)
	Not applicable	6 (12.8%)	27 (54.0%)	24 (63.2%)
	Not available	2	2	1
Breast cancer in history (invasive or <i>in situ</i>)	No	49 (100.0%)	44 (84.6%)	35 (89.7%)
	Yes	0 (0.0%)	8 (15.4%)	4 (10.3%)
	Not available	0	0	0

Table 2: Methylation analysis comparing nipple fluid of healthy controls and breast cancer patients

A) Univariable analysis for all 13 genes in the original panel																																																																																																																																						
Gene	Missing (%)	Methylation threshold (%) ^a	Methylation	Breast cancer (N)	Healthy control (N)	Univariable OR (95%CI)	P-value																																																																																																																															
<i>MGMT</i>	14	0.000	Positive	7	1	6.0 (0.7 – 52.1)	0.1																																																																																																																															
			Negative	45	48			<i>RARB</i>	14	0.583	Positive	12	3	4.5 (0.9 – 22.3)	0.07	Negative	40	46	<i>AKR1B1</i>	16	0.000	Positive	8	2	3.3 (0.6 – 16.6)	0.15	Negative	45 ^b	47	<i>ALX1</i>	16	1.127	Positive	11	5	2.4 (0.6 – 8.8)	0.2	Negative	41	44	<i>APC</i>	16	0.021	Positive	6	5	1.1 (0.2 – 5.3)	0.9	Negative	46	44	<i>RASSF1A</i>	18	0.046	Positive	8	3	2.7 (0.6 – 12.6)	0.22	Negative	44	46	<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15	Negative	45	48	<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3
<i>RARB</i>	14	0.583	Positive	12	3	4.5 (0.9 – 22.3)	0.07																																																																																																																															
			Negative	40	46			<i>AKR1B1</i>	16	0.000	Positive	8	2	3.3 (0.6 – 16.6)	0.15	Negative	45 ^b	47	<i>ALX1</i>	16	1.127	Positive	11	5	2.4 (0.6 – 8.8)	0.2	Negative	41	44	<i>APC</i>	16	0.021	Positive	6	5	1.1 (0.2 – 5.3)	0.9	Negative	46	44	<i>RASSF1A</i>	18	0.046	Positive	8	3	2.7 (0.6 – 12.6)	0.22	Negative	44	46	<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15	Negative	45	48	<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46						
<i>AKR1B1</i>	16	0.000	Positive	8	2	3.3 (0.6 – 16.6)	0.15																																																																																																																															
			Negative	45 ^b	47			<i>ALX1</i>	16	1.127	Positive	11	5	2.4 (0.6 – 8.8)	0.2	Negative	41	44	<i>APC</i>	16	0.021	Positive	6	5	1.1 (0.2 – 5.3)	0.9	Negative	46	44	<i>RASSF1A</i>	18	0.046	Positive	8	3	2.7 (0.6 – 12.6)	0.22	Negative	44	46	<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15	Negative	45	48	<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																	
<i>ALX1</i>	16	1.127	Positive	11	5	2.4 (0.6 – 8.8)	0.2																																																																																																																															
			Negative	41	44			<i>APC</i>	16	0.021	Positive	6	5	1.1 (0.2 – 5.3)	0.9	Negative	46	44	<i>RASSF1A</i>	18	0.046	Positive	8	3	2.7 (0.6 – 12.6)	0.22	Negative	44	46	<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15	Negative	45	48	<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																												
<i>APC</i>	16	0.021	Positive	6	5	1.1 (0.2 – 5.3)	0.9																																																																																																																															
			Negative	46	44			<i>RASSF1A</i>	18	0.046	Positive	8	3	2.7 (0.6 – 12.6)	0.22	Negative	44	46	<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15	Negative	45	48	<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																							
<i>RASSF1A</i>	18	0.046	Positive	8	3	2.7 (0.6 – 12.6)	0.22																																																																																																																															
			Negative	44	46			<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15	Negative	45	48	<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																		
<i>TM6SF1</i>	21	0.000	Positive	7	1	7.5 (0.5 – 118.4)	0.15																																																																																																																															
			Negative	45	48			<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32	Negative	43	44	<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																													
<i>CCND2</i>	26	0.003	Positive	9	5	2.2 (0.5 – 10.0)	0.32																																																																																																																															
			Negative	43	44			<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43	Negative	46	46	<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																																								
<i>SCGB3A1</i>	26	0.001	Positive	6	3	1.9 (0.4 – 9.7)	0.43																																																																																																																															
			Negative	46	46			<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21	Negative	45	47	<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																																																			
<i>GPX7</i>	34	0.000	Positive	7	2	3.9 (0.5 – 32.1)	0.21																																																																																																																															
			Negative	45	47			<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81	Negative	47 ^b	45	<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																																																														
<i>NDRG2</i>	35	0.140	Positive	6	4	1.2 (0.2 – 7.6)	0.81																																																																																																																															
			Negative	47 ^b	45			<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92	Negative	51	48	<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																																																																									
<i>MAL</i>	39	0.000	Positive	1	1	1.1 (0.1 – 19.1)	0.92																																																																																																																															
			Negative	51	48			<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91	Negative	48	46																																																																																																																				
<i>GSTP1</i>	43	0.134	Positive	4	3	1.1 (0.1 – 11.4)	0.91																																																																																																																															
			Negative	48	46																																																																																																																																	

^a Based on the 90th methylation percentile in healthy nipple fluid samples.

^b Numbers after digit were rounded to complete numbers, explaining that the sum of numbers is more than 52.

B) Multivariable analysis for <i>AKR1B1</i> , <i>ALX1</i> , <i>RASSF1A</i> and <i>TM6SF1</i>						
Gene	Methylated breast cancer (%)	Methylated healthy control (%)	Multivariable OR (95%CI)	P-value	AUC (95%CI)	P-value
<i>AKR1B1</i>	14	5	2.12 (0.37 – 12.05)	0.39	0.64 (0.54 – 0.74)	<0.01
<i>ALX1</i>	20	10	2.10 (0.54 – 8.22)	0.29		
<i>RASSF1A</i>	15	7	2.02 (0.38 – 10.64)	0.41		
<i>TM6SF1</i>	13	2	64.34 (0.00 – 3.67 10 ¹⁴)	0.78		

an AUC of 0.47 (95%CI 0.33 – 0.61, p=0.902). Using the CMI, AUC was 0.61 (95%CI 0.44 – 0.77, p=0.20), see also Figure 2F.

Methylation in nipple fluid of breast cancer patients and paired tumor tissue

In general the methylation levels in the tumor tissues were higher than in the nipple fluid samples. To evaluate correlation of methylation values in nipple fluid and corresponding tumor tissue, we calculated the percent of variation in methylation in tumor tissue explained by methylation in nipple fluid. For this analysis, all 52 breast cancer samples were used. Explained variance (R²) was 2.3%, 0.1%, 4.0% and 1.2% for *AKR1B1*, *ALX1*, *RASSF1A* and *TM6SF1*, respectively.

DISCUSSION

By evaluating methylation levels in nipple fluid samples obtained from healthy breasts versus affected and contralateral breasts from patients with breast cancer, this study aimed to investigate whether nipple fluid holds promise as a source of biomarkers for (early) breast cancer detection.

We demonstrated that nipple fluid can be obtained from breast cancer patients by aspiration under vacuum, supported by intranasal oxytocin, and that methylation analysis of these low-volume nipple fluid samples is feasible, given that a very sensitive PCR method such as QM-MSP is used [8]. Success rates of obtaining nipple fluid in breast cancer patients were in our experience lower than in healthy women or women at increased breast cancer risk [6, 7]. Since stress decreases lactation performance [13, 14], we expect that stress experienced following breast cancer diagnosis might partly explain the lower success rates in obtaining NAF. Possibly the success rates could be increased by performing more attempts to obtain nipple fluid.

One potential limitation of our study is the age difference between cases and controls possibly leading to methylation differences [15, 16]. In breast cancer, methylation was shown to correlate with age. However, higher age does not necessarily correlate with higher methylation percentages. With increasing age, methylation levels can both in- or decrease [17]. Consequently, only reports on methylation in breast tissue of the genes used in our panel could help to confirm that age-dependency of methylation is a confounder. DNA promoter methylation of a gene panel including *APC*, *CCND2*, *RARB*, *RASSF1A*,

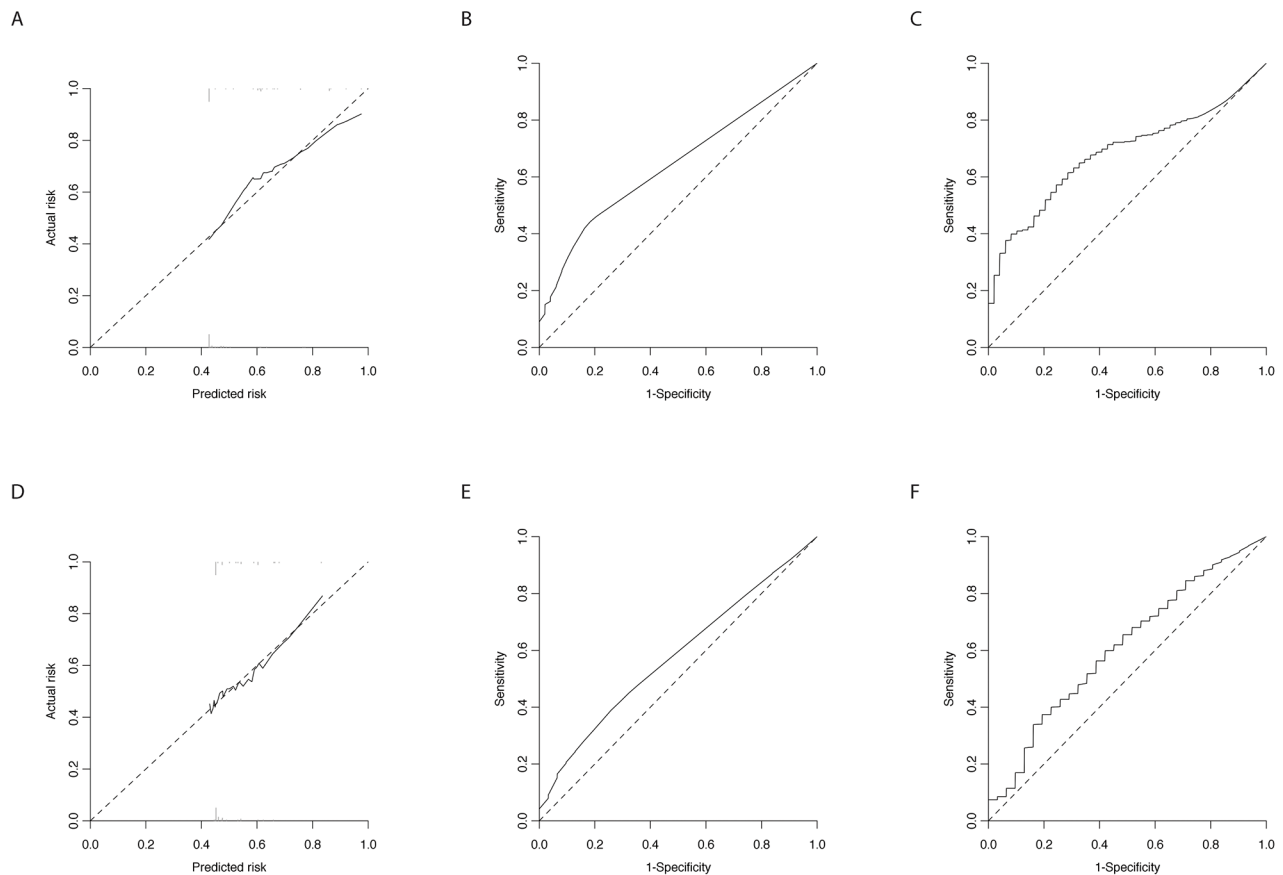


Figure 2: Calibration and ROC curves. Calibration A. and ROC B. curves of the 4 gene model (*AKR1B1*, *ALX1*, *RASSF1A* and *TM6SF1*) and the ROC C. curve for the CMI comparing methylation in nipple fluid from healthy controls to cancerous nipple fluid. Graphs D, E, and F. show the same curves for the analysis of nipple fluid from breasts with breast cancer and the contralateral counterpart.

Table 3: Methylation analysis comparing nipple fluid of breasts with cancer and their contralateral counterpart

A) Univariable analysis for all 13 genes in the original panel																																																																																																																																						
Gene	Missing (%)	Methylation threshold (%) ^a	Methylation	Affected side (N)	Contralateral side (N)	Univariable OR (95%CI)	P-value																																																																																																																															
<i>AKR1B1</i>	10	0.000	Positive	4	2	2.5 (0.3 – 21.3)	0.39																																																																																																																															
			Negative	27	29			<i>ALX1</i>	10	1.127	Positive	5	3	2.0 (0.3 – 11.6)	0.45	Negative	26	28	<i>RASSF1A</i>	11	0.046	Positive	4	3	1.5 (0.3 – 8.5)	0.64	Negative	27	28	<i>TM6SF1</i>	16	0.000	Positive	5	5	1.3 (0.3 – 5.5)	0.75	Negative	26	27 ^b	<i>CCND2</i>	18	0.003	Positive	7	4	2.1 (0.5 – 9.1)	0.35	Negative	24	27	<i>SCGB3A1</i>	18	0.001	Positive	5	3	2.3 (0.4 – 12.1)	0.34	Negative	26	28	<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1	Negative	25	25	<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1
<i>ALX1</i>	10	1.127	Positive	5	3	2.0 (0.3 – 11.6)	0.45																																																																																																																															
			Negative	26	28			<i>RASSF1A</i>	11	0.046	Positive	4	3	1.5 (0.3 – 8.5)	0.64	Negative	27	28	<i>TM6SF1</i>	16	0.000	Positive	5	5	1.3 (0.3 – 5.5)	0.75	Negative	26	27 ^b	<i>CCND2</i>	18	0.003	Positive	7	4	2.1 (0.5 – 9.1)	0.35	Negative	24	27	<i>SCGB3A1</i>	18	0.001	Positive	5	3	2.3 (0.4 – 12.1)	0.34	Negative	26	28	<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1	Negative	25	25	<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30						
<i>RASSF1A</i>	11	0.046	Positive	4	3	1.5 (0.3 – 8.5)	0.64																																																																																																																															
			Negative	27	28			<i>TM6SF1</i>	16	0.000	Positive	5	5	1.3 (0.3 – 5.5)	0.75	Negative	26	27 ^b	<i>CCND2</i>	18	0.003	Positive	7	4	2.1 (0.5 – 9.1)	0.35	Negative	24	27	<i>SCGB3A1</i>	18	0.001	Positive	5	3	2.3 (0.4 – 12.1)	0.34	Negative	26	28	<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1	Negative	25	25	<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																	
<i>TM6SF1</i>	16	0.000	Positive	5	5	1.3 (0.3 – 5.5)	0.75																																																																																																																															
			Negative	26	27 ^b			<i>CCND2</i>	18	0.003	Positive	7	4	2.1 (0.5 – 9.1)	0.35	Negative	24	27	<i>SCGB3A1</i>	18	0.001	Positive	5	3	2.3 (0.4 – 12.1)	0.34	Negative	26	28	<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1	Negative	25	25	<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																												
<i>CCND2</i>	18	0.003	Positive	7	4	2.1 (0.5 – 9.1)	0.35																																																																																																																															
			Negative	24	27			<i>SCGB3A1</i>	18	0.001	Positive	5	3	2.3 (0.4 – 12.1)	0.34	Negative	26	28	<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1	Negative	25	25	<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																							
<i>SCGB3A1</i>	18	0.001	Positive	5	3	2.3 (0.4 – 12.1)	0.34																																																																																																																															
			Negative	26	28			<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1	Negative	25	25	<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																		
<i>RARB</i>	18	0.583	Positive	6	6	1.0 (0.2 – 4.4)	1																																																																																																																															
			Negative	25	25			<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79	Negative	27	26	<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																													
<i>APC</i>	19	0.021	Positive	4	5	0.8 (0.2 – 4.2)	0.79																																																																																																																															
			Negative	27	26			<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87	Negative	28	28	<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																																								
<i>MGMT</i>	21	0.000	Positive	3	3	1.2 (0.2 – 9.3)	0.87																																																																																																																															
			Negative	28	28			<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92	Negative	31	31	<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																																																			
<i>MAL</i>	37	0.000	Positive	0	0	1.2 (0.0 – 64.8)	0.92																																																																																																																															
			Negative	31	31			<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52	Negative	28	26	<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																																																														
<i>NDRG2</i>	37	0.140	Positive	3	5	0.5 (0.1 – 3.6)	0.52																																																																																																																															
			Negative	28	26			<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17	Negative	25	30	<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																																																																									
<i>GPX7</i>	39	0.000	Positive	6	1	6.5 (0.4 – 95.1)	0.17																																																																																																																															
			Negative	25	30			<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65	Negative	29	30																																																																																																																				
<i>GSTP1</i>	45	0.134	Positive	2	1	2.2 (0.1 – 60.8)	0.65																																																																																																																															
			Negative	29	30																																																																																																																																	

^a Based on the 90th methylation percentile in healthy nipple fluid samples.

^b Numbers after digit were rounded to complete numbers, explaining that the sum of numbers is more than 31.

B) Multivariable analysis for <i>AKR1B1</i> , <i>ALX1</i> , <i>RASSF1A</i> and <i>TM6SF1</i>						
Gene	Methylated affected side (%)	Methylated contralateral side (%)	Multivariable OR (95%CI)	P-value	AUC (95%CI)	P-value
<i>AKR1B1</i>	13	6	2.42 (0.27 – 21.65)	0.43	0.58 (0.44 – 0.72)	0.26
<i>ALX1</i>	16	9	1.86 (0.30 – 11.65)	0.51		
<i>RASSF1A</i>	14	10	1.43 (0.24 – 8.61)	0.69		
<i>TM6SF1</i>	18	15	1.09 (0.23 – 5.13)	0.91		

and *SCGB3A1* in breast cancer samples was not correlated with age in a previous study [18]. Similar results were shown in normal breast tissue, in which methylation of *RARB*, *RASSF1A*, and *SCGB3A1* hardly correlated to age. In cancer tissue, *SCGB3A1* was only weakly correlated to age, whereas the other genes were not [19]. We previously reported that the CMI of an 11-gene panel (*RARB*, *RASSF1*, *TWIST1*, *CCND2*, *ESR1*, *SCGB3A1*, *BRCA1*, *BRCA2*, *CDKN2A*, *APC*, *CDHI*) in breast tumors increased with age. However, age-dependency was not determined for the individual genes. When adjusted for age, CMI and the presence of malignancy were still associated [9]. Another complicating factor is that breast cancer risk also increases with age. As a result, it is challenging to discriminate if increasing methylation is due to ageing per se or to the increased breast cancer risk. In normal breast tissue, age-related methylation changes were further altered in breast tumors and may therefore represent early events contributing to breast carcinogenesis [20, 21]. In summary, it is not possible to make definitive assumptions about how age might influence methylation status of our candidate gene panel in nipple fluid or breast tissue. The best way to handle this problem would be to age-match the participants in our nipple fluid studies, but the current sample size does not allow this. Moreover, age is not the only factor possibly influencing methylation status. For example, obesity is associated with methylation in ER-positive tumors [22], making it difficult to ever obtain clinically equal groups of women only differing in breast cancer status.

Even though methylation values in nipple fluid were low, we could discriminate cancerous from healthy nipple fluid samples with an AUC of 0.68 using the CMI, and an AUC of 0.64 using a 4-gene panel. The latter was also

predictive without prior data imputation. The differences between cancerous and contralateral nipple fluid samples were less prominent, suggesting that the contralateral breast may undergo field effects and therefore making it more difficult to discriminate between the affected and healthy breast. This is in line with a previous proteomics study that demonstrates a similar protein expression profile in the affected and contralateral breast, but a distinct protein expression profile in healthy breasts [23]. Also for methylation, field effects have been described. In a set of six breast cancer patients, methylation of *RASSF1A* could be found up to 4 cm away from the tumor. Unfortunately, tissue further away was not investigated, but extensive methylation was found in the contralateral breast in two patients [24].

To further validate the critical role of our final 4 gene panel in breast carcinogenesis, we used genomic locations identified in previous studies [10, 11], data generated using the Illumina Human Methylation 27 Beadchip Array, and validated this in The Cancer Genome Atlas (TCGA) Breast Cancer Invasive Carcinoma data (<http://cancergenome.nih.gov/>). QM-MSP primers were designed to overlap or hybridize to a region within 100 bp of the array probe genomic location. With a methylation threshold of 15%, the frequency of methylation in TCGA breast cancers (N=312) was 64.4% for *AKR1B1*, 67.3% for *ALX1*, 81.4% for *RASSF1A*, and 52.2% for *TM6SF1*. These numbers again stress the importance of methylation of the 4 selected genes/CpG regions in breast carcinogenesis.

Table 4 gives an overview of previous reports describing methylation analysis of nipple aspirate or ductal lavage fluid and demonstrates variable results regarding the diagnostic value of methylation analysis in

Table 4: Literature overview of methylation analysis in ductal lavage/nipple fluid

Author, year	Source of fluid	No. of cases	No. of controls	Used assay	Genes tested	Results
Krassenstein, 2004 [43]	Nipple	22 BC patients	5 healthy and 5 benign disease	MSP	<i>GSTP1</i> , <i>RAB2</i> , <i>CDKN2A</i> (p16INK4a, p14ARF), <i>RASSF1A</i> , <i>DAPK</i>	Once the individual tumor is positive for hypermethylation of a candidate gene, corresponding fluid was analysed for hypermethylation of that particular gene. Hypermethylation of one or more genes was found in all tumor DNAs and identical gene hypermethylation was detected in 82% matched nipple fluid. Hypermethylation was absent in healthy women.
Fackler, 2006 [8]	Ductal	27 BC patients	3	QM-MSP	<i>RASSF1A</i> , <i>TWIST</i> , <i>SCGB3A1</i> , <i>CCND2</i> , <i>RARB</i> , <i>APC1</i> , <i>BRCA1</i> , <i>BRCA2</i> , <i>CDKN2A</i> (p16INK4a)	Sensitivity of 62% and specificity of 82% for detecting breast cancer by CMI.

(Continued)

Author, year	Source of fluid	No. of cases	No. of controls	Used assay	Genes tested	Results
Euhus, 2007 [44]	Ductal	67 BC patients	83	QM-MSP	<i>CCND2, APC, SCGB3A1, RASSF1A, RARB2</i>	QM-MSP data available for 34 tumor-tissue FNAs from patients participating in the ductal lavage study. Frequency of methylation was significantly higher in tumors than corresponding fluid for every gene. Among methylation-positive cases, median methylation fraction was significantly increased for tumor tissue than corresponding fluid for <i>APC, HIN1</i> , and <i>RASSF1A</i> . There was no correlation between individual methylation fractions of tumors as compared to corresponding fluid for any gene.
Locke, 2007 [45]	Ductal	7 <i>BRCA1</i> , 12 <i>BRCA2</i> mutation carriers	5	MSP	<i>RARB, SCGB3A1, TWIST, CCND2</i>	In 42% of <i>BRCA</i> mutation carriers, at least one gene was hypermethylated, but no hypermethylation was observed control patients (p=0.13). Methylation was detected in 61% of <i>BRCA</i> mutation carriers. Methylation of <i>CDKN2A</i> (p16INK4a) was observed in 37% that coincided with a <i>BRCA1</i> mutation, <i>RASSF1A</i> methylation in 24%, <i>TWIST</i> methylation in 6% that was associated with cytologic atypia, and <i>RARB</i> methylation in 24% that coincided with an association between NAF production and methylation in this gene at individual breast level.
Antill, 2010 [46]	Ductal	16 <i>BRCA1</i> , 18 <i>BRCA2</i> mutation carriers; 7 developing BC (1 bilateral)	None	QM-MSP	<i>CDKN2A</i> (p16INK4a), <i>RASSF1A, TWIST, RARB</i>	NAF versus normal tissue showed only a significant difference in hypermethylation of <i>CDKN2A</i> (p16INK4a); significant difference in cancer compared to benign ductal lavage for p16 and <i>RASSF1</i> . No significant difference between cancer and precancer samples as well as between precancer and benign samples.
Zhu, 2010 [47]	Nipple; Ductal	18; 61 BC and 10 precancer	None; 48 benign disease	Q-MSP	<i>CCND2, CDKN2A</i> (p16INK4a), <i>RARB, RASSF1A</i>	Significant difference in hypermethylation of <i>SCGB3A1, CDH13, RARB</i> and <i>IGFBP7</i> between breast cancer and healthy samples. ROC 0.76, false positive 33%, false negative 16%.
Twelves, 2013 [48]	Ductal	54	46 contralateral	MSP	<i>RASSF1A, SCGB3A1, PDLIM4, CDH13, RARB, IGFBP7</i>	

BC = breast cancer; NAF = nipple aspiratin fluid

locally derived breast fluid. Hence, the extent to which the low sample volume and the possible dilution with normal epithelial cells could have contributed to the limited diagnostic accuracy is unclear at this point, and needs to be explored further.

Apart from breast fluids, methylation as a biomarker has been studied in different other types of body fluids, such as vaginal swabs [25], urine [26], or stool samples [27], generally yielding much better results. Moreover, studies with other, high-volume biofluids show a better correlation between fluid and tumor tissue, e.g. in urine [28] and stool [29]. The larger volume of the samples in these studies compared to nipple fluid may have contributed to the higher reported diagnostic accuracies. Moreover, vaginal swabs and urine provide a more direct way of sampling from the source organ, whereas usually six to eight ducts reach the human nipple and it remains elusive whether shedded epithelial cells and DNA from the cancerous duct are efficiently collected in the obtained nipple fluid. Although there is evidence for a methylation field defect in breast cancer, nipple fluid samples may be diluted with non-diseased material. This could account for the low correlation observed between methylation in nipple fluid and the corresponding breast cancer tissue. Collecting fluid via ductal lavage might improve diagnostic accuracy by providing a larger sample volume and allowing more direct sampling of the affected ducts, at the costs of a more invasive procedure being less applicable in a screening setting. In addition, other classes of biomarkers may improve diagnostic accuracy such as proteins [30-32], hormones [33-35], lipids [36] or microRNAs.

In conclusion, we have clearly demonstrated that cancerous nipple fluid contains increased levels of methylation biomarkers that may help to detect breast cancer in a non-invasive way. As part of a large prospective trial in cooperation with the Erasmus Medical Center, the Netherlands, we are currently yearly collecting nipple fluid of women at increased breast cancer risk for 10 years, resulting in a valuable biobank of nipple fluid samples. In the future, this biobank will allow us to test the predictive value of nipple fluid biomarkers in a prospective setting, so appropriate preventive measures can be taken.

MATERIALS AND METHODS

Nipple fluid and tissue samples

The clinical study collecting nipple fluid from healthy women by aspiration under vacuum, supported by intranasal oxytocin, was described previously [6]. Similarly, we collected nipple fluid from sporadic breast cancer patients using a cross-sectional study design between 2010 and 2014. Women were included when having proven or suspected breast cancer based on biopsy, either invasive or *in situ*. Exclusion criteria were age under 18 years, bilateral ablative breast surgery,

pregnancy or lactation, having an active breast infection and/or having disseminated breast cancer. Nipple fluid was collected prior to breast cancer surgery from the affected and contralateral breast as described before [6]. TNE buffer (50 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA) was added and the nipple fluid sample was stored at -80 °C until further analysis blinded for breast cancer status. Formalin-fixed paraffin-embedded (FFPE) breast cancer tissue, collected at the time of surgery, was used after diagnostics had been completed. The study was approved by the Internal Review Board of the UMC Utrecht, the Netherlands (ABR NL 11690.041.06, METC 06-091). Written informed consent was obtained from all participants.

DNA extraction

DNA was isolated from nipple fluid samples using the High Pure Viral Nucleic Acid kit (Roche, 11858874001) and low-retention Eppendorf tips (Biotix, 2012091). Preferably 10 µl of the sample was used for DNA isolation, and DNA was isolated twice for duplicate analysis. If less than the required amount was available, all nipple fluid was used.

For DNA isolation from FFPE tissue, one to five 10 µm unstained sections were deparaffinized in xylene and rehydrated through a series of alcohol. Relevant tissue, as indicated by a pathologist on sandwich H&E stained sections, was scraped from the slide and 100 µl lysis buffer (0.5% Tween-20, 50 mM Tris pH 8.5) containing 20-40 µg proteinase K (Invitrogen) was added. After incubation at 56°C overnight, the reaction was heat inactivated for 10 min at 95°C and centrifuged at 14,000 rpm for 3 min. The supernatant was transferred to a new tube and DNA concentration as well as 260/280 absorbance were measured with a spectrophotometer (NanoDrop ND-1000, Thermo Scientific). Samples were stored at 4°C until further analysis.

Sodium bisulfite conversion

Sodium bisulfite conversion was performed using the Epitect bisulfite kit (Qiagen, 59104) according to the protocol "Sodium Bisulfite Conversion of Unmethylated Cytosines in Small Amounts of Fragmented DNA" (protocol version 09/2009) with an input of 40 µl of DNA from nipple fluid and 1.5 µg DNA from tissue. Human sperm DNA was used as a negative control and *SssI* methylase treated MDA-MB-231 gDNA as a positive control. A non-template control and 3% methylated control were included in each bisulfite conversion reaction.

External multiplex PCR

Immediately after sodium bisulfite treatment, multiplex PCR was performed with 12 µl of converted DNA kept on ice and a primer mix to amplify the

following 13 genes: *AKR1B1*, *ALX1*, *APC1*, *CCND2*, *GPX7*, *GSTP1*, *HIN1*, *MAL*, *MGMT*, *NDRG2*, *RARB*, *RASSF1A*, and *TM6SF1*. This set of genes was chosen after validation in breast cancer tissue [9-12] and based on their initial performance in nipple fluid samples. Primer sequences are shown in Supplementary Table S2. For each PCR reaction, converted DNA was added to 13 μ l of reaction mix consisting of 1x MSP buffer (67 mM Tris (pH 8.8), 6.7 mM $MgCl_2$, 10 mM β -mercaptoethanol, 0.1% DMSO), 1.25 mM dNTPs, 2.5 units of Platinum Taq (Invitrogen) and 0.1 μ M each of the forward and reverse primers (dissolved in distilled water containing 50 μ g/ml Salmon Sperm DNA). PCR conditions were 95°C for 5 minutes, followed by 40 cycles of 95°C for 30 seconds, 56°C for 45 seconds and 72°C for 45 seconds, with a final extension cycle of 72°C for 7 minutes. PCR products were diluted, 500x for NAF and 500-5000x for tissue, in dilution buffer (distilled water containing 1x MSP buffer and 100 μ g/ml Salmon Sperm DNA) for further analysis.

Gene-specific quantitative PCR (Q-PCR)

To investigate the specific target gene of interest, a nested real-time methylation specific PCR (QM-MSP) [8] was performed for each candidate gene separately. Used primer and probe sequences are listed in Supplementary Tables S3 and S4. For each reaction, 4 μ l of diluted multiplex PCR product was added to 16 μ l reaction mix consisting of 1x MSP buffer, 200 μ M dNTPs, 500 nM ROX 50 (Invitrogen), 1 unit of Platinum Taq (Invitrogen), 400 nM of each primer, and 400 nM of each probe. Q-PCR was performed on the Applied Biosystems ViiA7 Real-Time PCR System with the following conditions: 50°C for 2 minutes, 95°C for 7 minutes, followed by 50 cycles of 95°C for 15 seconds and 65°C for 1 minute. For each gene, a standard dilution curve (dilution 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} , 10^{-7}) of a mixed sample containing fully methylated *SssI* methylase treated MDA-MB-231 gDNA and unmethylated human sperm DNA in a 1:1 ratio was included for extrapolating percent methylation. Analysis was performed blinded for tumor status.

Statistical analysis

As methylation status could not be assessed for all genes in all samples, and since disregarding missing data may lead to biased results and loss of information [37, 38], we imputed missing data by the multivariate imputation by chained equations (MICE) method (20 datasets) [39]. We used extensive information for imputing missing data: baseline characteristics including reproductive factors, cancer status of the breast, tumor characteristics if applicable and methylation percentages (Table 1, Supplementary Table S1). All analyses were performed in each imputed dataset separately and pooled using Rubin's rules.

After imputation, we first compared methylation between healthy volunteers and patients with breast cancer. For this, we used nipple fluid samples collected unilaterally in healthy volunteers (N=49) and nipple fluid from breast cancer patients (N=52; affected side). In case of two samples from the same woman in the control group, we randomly selected one of the samples. Second, methylation between the affected and paired contralateral breasts was compared (N=31). Finally, methylation in nipple fluid was compared to methylation status in tumor tissue of the same breast (N=52).

To compare methylation values of the nipple fluid samples, methylation signals were evaluated as a dichotomous variable, based on the 90th percentile in the healthy control group (N=88). The threshold of the 90th percentile was based on a study analysing methylation in ductal fluid [8]. To begin with, we evaluated the discriminating value of each gene in a univariable comparison. For multivariable logistic regression analysis, we selected 4 genes, since less than 10 events per variable statistical models may not be valid [40, 41]. We chose the 4 genes with the lowest amount of missing data in the bilateral sample set (N=31). We used the same panel of genes in the comparison of healthy versus breast cancer samples, in order to make all analyses comparable although this analysis formally allowed for more genes to be evaluated.

Next, we evaluated the predictive performance of each logistic regression model for discrimination and calibration. Discrimination was assessed with the area under the receiver operating characteristic (ROC) curve (AUC). To evaluate how close the model-derived predicted cancer probabilities reflected observed probabilities over the entire range of possible values, we made use of calibration plots. To validate the results of the 4-gene panel in non-imputed data, we performed multivariable logistic regression and ROC analysis of the cases with non-missing methylation values for *ALX1*, *AKR1B1*, *RASSF1*, and *TM6SF1* genes. The number of complete cases was 29, 40 and 28 for healthy, cancerous and contralateral nipple fluid samples, respectively.

In addition to using dichotomous variables, we also calculated the sum of the methylation percentages of all 13 analysed genes, the cumulative methylation index (CMI) [42]. Also for CMI, an AUC was calculated.

To evaluate correlation of methylation values in nipple fluid and corresponding tumor tissue, we calculated the percent of variation in methylation in tumor tissue explained by methylation in nipple fluid.

Statistical analyses were performed using IBM SPSS Statistics version 20.0 (SPSS, Inc., Chicago, IL, USA) and R version 3.1.1 (The R foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>) including package Regression Modeling Strategies (<http://CRAN.R-project.org/package=rms>). We considered a two-sided p-value <0.05 as statistically significant, and report parameter estimates with 95% confidence intervals (CI).

ACKNOWLEDGMENTS

We would like to thank N. Dijkstra, A. Hennink, H. Meijrink, and B. Verolme for collecting the nipple fluid samples and T. van Dalen for help with the inclusion of patients. We would like to acknowledge J. Meeldijk and X. Pan for optimizing the QM-MSP technique.

COMPETING INTERESTS

The author(s) declare that they have no competing interests.

FINANCIAL SUPPORT

This work was supported by Aegon, A Sister's Hope [11.076], Dutch Cancer Society [UU 2008-4217], Pink Ribbon [0.8033 via Stichting Vrienden UMC Utrecht], and ZonMW [92003506].

REFERENCES

1. GLOBOCAN 2008. Most frequent cancers.
2. Integraal Kankercentrum Nederland. Mammacarcinoom, Landelijke richtlijn, Versie: 2.0. 2012.
3. Suijkerbuijk KP, van der Wall E, Vooijs M, van Diest PJ. Molecular analysis of nipple fluid for breast cancer screening. *Pathobiology*. 2008; 75: 149-152.
4. Suijkerbuijk KP, van Diest PJ, van der Wall E. Improving early breast cancer detection: focus on methylation. *Annals of Oncology*. 2011; 22: 24-29.
5. Fabian CJ, Kimler BF, Mayo MS, Khan SA. Breast-tissue sampling for risk assessment and prevention. *Endocrine-related cancer*. 2005; 12: 185-213.
6. Suijkerbuijk KP, van der Wall E, van Diest PJ. Oxytocin: bringing magic into nipple aspiration. *Annals of Oncology*. 2007; 18: 1743-1744.
7. Suijkerbuijk KP, van der Wall E, Meijrink H, Pan X, Borel Rinkes IH, Ausems MG, van Diest PJ. Successful oxytocin-assisted nipple aspiration in women at increased risk for breast cancer. *Familial cancer*. 2010; 9: 321-325.
8. Fackler MJ, Malone K, Zhang Z, Schilling E, Garrett-Mayer E, Swift-Scanlan T, Lange J, Nayar R, Davidson NE, Khan SA, Sukumar S. Quantitative multiplex methylation-specific PCR analysis doubles detection of tumor cells in breast ductal fluid. *Clinical cancer research*. 2006; 12: 3306-3310.
9. Suijkerbuijk KP, Fackler MJ, Sukumar S, van Gils CH, van Laar T, van der Wall E, Vooijs M, van Diest PJ. Methylation is less abundant in BRCA1-associated compared with sporadic breast cancer. *Annals of Oncology*. 2008; 19: 1870-1874.
10. Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvanathan K, Marks J, Ethier S, Gray JW, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer research*. 2011; 71: 6195-6207.
11. Fackler MJ, Lopez Bujanda Z, Umbricht C, Teo WW, Cho S, Zhang Z, Visvanathan K, Jeter S, Argani P, Wang C, Lyman JP, de Brot M, Ingle JN, et al. Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. *Cancer research*. 2014; 74: 2160-2170.
12. de Groot JS, Pan X, Meeldijk J, van der Wall E, van Diest PJ, Moelans CB. Validation of DNA promoter hypermethylation biomarkers in breast cancer--a short report. *Cellular oncology (Dordrecht)*. 2014; 37: 297-303.
13. Chen DC, Nommsen-Rivers L, Dewey KG, Lonnerdal B. Stress during labor and delivery and early lactation performance. *The American Journal of Clinical Nutrition*. 1998; 68: 335-344.
14. Dewey KG. Maternal and fetal stress are associated with impaired lactogenesis in humans. *The Journal of nutrition*. 2001; 131: 3012S-5S.
15. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature genetics*. 1994; 7: 536-540.
16. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer research*. 1998; 58: 5489-5494.
17. Euhus DM, Bu D, Milchgrub S, Xie XJ, Bian A, Leitch AM, Lewis CM. DNA methylation in benign breast epithelium in relation to age and breast cancer risk. *Cancer epidemiology, biomarkers & prevention*. 2008; 17: 1051-1059.
18. Swift-Scanlan T, Vang R, Blackford A, Fackler MJ, Sukumar S. Methylated genes in breast cancer: associations with clinical and histopathological features in a familial breast cancer cohort. *Cancer biology & therapy*. 2011; 11: 853-865.
19. Feng W, Shen L, Wen S, Rosen DG, Jelinek J, Hu X, Huan S, Huang M, Liu J, Sahin AA, Hunt KK, Bast RC Jr, Shen Y, et al. Correlation between CpG methylation profiles and hormone receptor status in breast cancers. *Breast cancer research*. 2007; 9: R57.
20. Johnson KC, Koestler DC, Cheng C, Christensen BC. Age-related DNA methylation in normal breast tissue and its relationship with invasive breast tumor methylation. *Epigenetics*. 2014; 9: 268-275.
21. Xu Z, Taylor JA. Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis*. 2014; 35: 356-364.
22. Hair BY, Troester MA, Edmiston SN, Parrish EA, Robinson WR, Wu MC, Olshan AF, Swift-Scanlan T, Conway K. Body mass index is associated with gene methylation in estrogen receptor-positive breast tumors. *Cancer epidemiology, biomarkers & prevention*. 2015; 24: 580-586.

23. Noble JL, Dua RS, Coulton GR, Isacke CM, Gui GP. A comparative proteomic analysis of nipple aspiration fluid from healthy women and women with breast cancer. *European journal of cancer (Oxford, England : 1990)*. 2007; 43: 2315-2320.
24. Yan PS, Venkataramu C, Ibrahim A, Liu JC, Shen RZ, Diaz NM, Centeno B, Weber F, Leu YW, Shapiro CL, Eng C, Yeatman TJ, Huang TH. Mapping geographic zones of cancer risk with epigenetic biomarkers in normal breast tissue. *Clinical cancer research*. 2006; 12: 6626-6636.
25. Doufekas K, Hadwin R, Kandimalla R, Jones A, Mould T, Crowe S, Olaitan A, Macdonald N, Fiegl H, Wik E, Salvesen HB, Widschwendter M. GALR1 methylation in vaginal swabs is highly accurate in identifying women with endometrial cancer. *International journal of gynecological cancer*. 2013; 23: 1050-1055.
26. Monteiro-Reis S, Leca L, Almeida M, Antunes L, Monteiro P, Dias PC, Morais A, Oliveira J, Henrique R, Jeronimo C. Accurate detection of upper tract urothelial carcinoma in tissue and urine by means of quantitative GDF15, TMEFF2 and VIM promoter methylation. *European journal of cancer (Oxford, England : 1990)*. 2014; 50: 226-233.
27. Bosch LJ, Oort FA, Neerincx M, Khalid-de Bakker CA, Terhaar sive Droste JS, Melotte V, Jonkers DM, Masclee AA, Mongera S, Grooteclaes M, Louwagie J, van Criekinge W, Coupe VM, et al. DNA methylation of phosphatase and actin regulator 3 detects colorectal cancer in stool and complements FIT. *Cancer prevention research (Philadelphia, Pa.)*. 2012; 5: 464-472.
28. Berrada N, Amzazi S, Ameziane El Hassani R, Benbacer L, El Mzibri M, Khyatti M, Chafiki J, Abbar M, Al Bouzidi A, Ameer A, Attaleb M. Epigenetic alterations of adenomatous polyposis coli (APC), retinoic acid receptor beta (RARbeta) and survivin genes in tumor tissues and voided urine of bladder cancer patients. *Cellular and molecular biology (Noisy-le-Grand, France)*. 2012; Suppl.58: OL1744-51.
29. Xiao Z, Li B, Wang G, Zhu W, Wang Z, Lin J, Xu A, Wang X. Validation of methylation-sensitive high-resolution melting (MS-HRM) for the detection of stool DNA methylation in colorectal neoplasms. *Clinica chimica acta*. 2014; 431: 154-163.
30. Sauter ER, Shan S, Hewett JE, Speckman P, Du Bois GC. Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. *International journal of cancer*. 2005; 114: 791-796.
31. Pawlik TM, Fritsche H, Coombes KR, Xiao L, Krishnamurthy S, Hunt KK, Pusztai L, Chen JN, Clarke CH, Arun B, Hung MC, Kuerer HM. Significant differences in nipple aspirate fluid protein expression between healthy women and those with breast cancer demonstrated by time-of-flight mass spectrometry. *Breast cancer research and treatment*. 2005; 89: 149-157.
32. Mannello F, Qin W, Zhu W, Fabbri L, Tonti GA, Sauter ER. Nipple aspirate fluids from women with breast cancer contain increased levels of group IIA secretory phospholipase A2. *Breast cancer research and treatment*. 2008; 111: 209-218.
33. Mannello F, Tonti GA, Pagliarani S, Benedetti S, Canestrari F, Zhu W, Qin W, Sauter ER. The 8-epimer of prostaglandin F(2alpha), a marker of lipid peroxidation and oxidative stress, is decreased in the nipple aspirate fluid of women with breast cancer. *International journal of cancer*. 2007; 120: 1971-1976.
34. Chatterton RT, Jr, Khan SA, Heinz R, Ivancic D, Lee O. Patterns of sex steroid hormones in nipple aspirate fluid during the menstrual cycle and after menopause in relation to serum concentrations. *Cancer epidemiology, biomarkers & prevention*. 2010; 19: 275-279.
35. Maskarinec G, Ollberding NJ, Conroy SM, Morimoto Y, Pagano IS, Franke AA, Gentzsch E, Stanczyk FZ. Estrogen levels in nipple aspirate fluid and serum during a randomized soy trial. *Cancer epidemiology, biomarkers & prevention*. 2011; 20: 1815-1821.
36. Djuric Z, Chen G, Ren J, Venkatramanamoorthy R, Covington CY, Kucuk O, Heilbrun LK. Effects of high fruit-vegetable and/or low-fat intervention on breast nipple aspirate fluid micronutrient levels. *Cancer epidemiology, biomarkers & prevention*. 2007; 16: 1393-1399.
37. Janssen KJ, Vergouwe Y, Donders AR, Harrell FE, Jr, Chen Q, Grobbee DE, Moons KG. Dealing with missing predictor values when applying clinical prediction models. *Clinical chemistry*. 2009; 55: 994-1001.
38. Janssen KJ, Donders AR, Harrell FE, Jr, Vergouwe Y, Chen Q, Grobbee DE, Moons KG. Missing covariate data in medical research: to impute is better than to ignore. *Journal of clinical epidemiology*. 2010; 63: 721-727.
39. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of statistical software*. 2011; 45: 1-67.
40. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *Journal of clinical epidemiology*. 1995; 48: 1503-1510.
41. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*. 1996; 49: 1373-1379.
42. Fackler MJ, Rivers A, Teo WW, Mangat A, Taylor E, Zhang Z, Goodman S, Argani P, Nayar R, Susnik B, Sukumar S, Khan SA. Hypermethylated genes as biomarkers of cancer in women with pathologic nipple discharge. *Clinical cancer research*. 2009; 15: 3802-3811.
43. Krassenstein R, Sauter E, Dulaimi E, Battagli C, Ehya H, Klein-Szanto A, Cairns P. Detection of breast cancer in nipple aspirate fluid by CpG island hypermethylation. *Clinical cancer research*. 2004; 10: 28-32.
44. Euhus DM, Bu D, Ashfaq R, Xie XJ, Bian A, Leitch AM, Lewis CM. Atypia and DNA methylation in nipple

- duct lavage in relation to predicted breast cancer risk. *Cancer epidemiology, biomarkers & prevention*. 2007; 16: 1812-1821.
45. Locke I, Kote-Jarai Z, Fackler MJ, Bancroft E, Osin P, Nerurkar A, Izatt L, Pichert G, Gui GP, Eeles RA. Gene promoter hypermethylation in ductal lavage fluid from healthy BRCA gene mutation carriers and mutation-negative controls. *Breast cancer research*. 2007; 9: R20.
46. Antill YC, Mitchell G, Johnson SA, Devereux L, Milner A, Di Iulio J, Lindeman GJ, Kirk J, Phillips KA, Campbell IG. Gene methylation in breast ductal fluid from BRCA1 and BRCA2 mutation carriers. *Cancer epidemiology, biomarkers & prevention*. 2010; 19: 265-274.
47. Zhu W, Qin W, Hewett JE, Sauter ER. Quantitative evaluation of DNA hypermethylation in malignant and benign breast tissue and fluids. *International journal of cancer*. 2010; 126: 474-482.
48. Twelves D, Nerurkar A, Osin P, Dexter T, Ward A, Gui GP, Isacke CM. DNA promoter hypermethylation profiles in breast duct fluid. *Breast cancer research and treatment*. 2013; 139: 341-350.