Hypermethylated Genes as Biomarkers of Cancer in Women with Pathologic Nipple Discharge

Mary Jo Fackler,[1] Aeisha Rivers,[3] Wei Wen Teo,[1] Amrit Mangat,[3] Evangeline Taylor,[1] Zhe Zhang,[1] Steve Goodman,[1] Pedram Argani,[2] Ritu Nayar,[4] Barbara Susnik,[4] Saraswati Sukumar,[1] and Seema A. Khan[4]

Abstract

Purpose: In a pilot study of women with pathologic nipple discharge (PND) undergoing ductoscopy, we tested quantitative assessment of gene promoter hypermethylation using quantitative multiplex methylation–specific PCR (QM-MSP) to enhance detection of duct carcinoma *in situ* (DCIS).

Experimental Design: Women with PND underwent ductoscopy; ducts with significant lesions were surgically resected (36 ducts in 33 women) and those with minimal findings were not (28 ducts in 16 women). QM-MSP was done on ductoscopy cell samples. Results were compared with cytology and tissue histology.

Results: Cells from ducts with significant lesions on ductoscopy had significantly higher levels of methylation than those with minimal findings. Furthermore, cells from ducts with DCIS displayed higher levels of methylation than those with benign lesions such as papilloma ($P = 0.006$); or ducts with minimal findings on ductoscopy ($P = 0.0001$). Cumulative *RASSF1A, TWIST1*, and *HIN1* gene methylation accurately distinguished ducts with cancerous versus benign lesions (100% sensitivity, 72% specificity, and area under the curve of 0.91 according to receiving operating characteristic analyses). QM-MSP analysis was more informative than cytology (100% versus 29% sensitivity, respectively), for detecting DCIS. In a validation set of paraffin-embedded DCIS and papilloma samples from women presenting with PND, QM-MSP was significantly higher in DNA from DCIS than papilloma sections ($P = 0.002$).

Conclusion: The positive predictive value of ductoscopy was more than doubled (19% versus 47%) with the addition of QM-MSP, demonstrating the benefit of targeting ducts having both high methylation and significant abnormalities on ductoscopy for surgical excision. Future large-scale studies to validate this approach are needed.

Pathologic nipple discharge (PND) contributes to 5% of referrals to breast surgeons in the western world (1, 2). The most frequent causes of PND in these cases are intraductal papilloma in 36% to 66% (2-7), duct carcinoma *in situ* (DCIS) in 3% to 20% (2-6, 8), and other benign causes in up to 23% (3, 4). The evaluation of women with PND usually involves mammography, ultrasonography, and frequently, ductography (galactography). Cytologic examination of PND is of variable utility, and is not uniformly done (4, 9). Recently, magnetic resonance imaging of the breast has been added to the diagnostic algorithm in some institutions (10). Each of these procedures has low sensitivity for detection of cancer, and often low specificity (3, 10, 11). Mammary endoscopy (ductoscopy) was first introduced in Japan in 1988 for the evaluation of patients with PND, and is currently being used to improve localization of lesions in patients with PND; it also allows retrieval of intraductal cells for diagnostic purposes (5, 12-17). Although useful for lesion localization, ductoscopy does not distinguish benign from malignant growth (6, 7, 18). Thus, PND is one of the few remaining breast problems where surgical excision is required for definitive diagnosis. The development of a nonsurgical method to reliably diagnose cancer would offer the possibility of diagnosis without surgery in women with PND who have a very low likelihood of significant neoplasia. In the future, a molecular diagnostic test that allows reliable diagnosis of intraductal

Authors' Affiliations: Departments of [1]Oncology and [2]Pathology Johns Hopkins University School of Medicine, Baltimore, Maryland; and Departments of [3]Surgery and [4]Pathology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois

Received 8/26/08; revised 2/16/09; accepted 2/20/09; published OnlineFirst 5/26/09.

Grant support: National Cancer Institute Specialized Programs of Research Excellence P50 CA88843 (S. Sukumar), the AVON/National Cancer Institute-PFP award CA88843-AV-14P1 (S. Sukumar and S.A. Khan), and Early Detection Research Network Associate Member grant (S.A. Khan).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Seema A. Khan, Robert H. Lurie Comprehensive Cancer Center of Northwestern University 303, East Superior Street, Lurie 4-111, Chicago, IL 60611. Phone 312-503-2880; Fax 312-503-2555; E-mail: skhan@nmh.org or Saraswati Sukumar, Breast Cancer Program Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, 1650 Orleans Street, Rm 410, Baltimore, MD. Phone: 410-614-2479; Fax: 410-614-4073; E-mail: saras@jhmi.edu.

doi:10.1158/1078-0432.CCR-08-1981

Translational Relevance

The work we describe has clear translational relevance. It represents the application of a well-validated laboratory observation (gene promoter methylation is a consistent event in malignant cells) to a clinical need [the ability to distinguish women with benign causes of nipple discharge from those with duct carcinoma *in situ* (DCIS)]. The methylation panel we have studied is the result of a careful series of experiments documented in past publications (see manuscript for references). In this pilot study, we show that the relatively new clinical technique of ductoscopy combined with the measurement of DNA methylation in three genes (*RASSF1A, HIN1*, and *TWIST1*) in the ductoscopy washings increases the classification accuracy in these patients by >2-fold compared with ductoscopy and cytologic evaluation. In a separate validation set of paraffin-embedded archival samples of DCIS, papilloma, and normal breast tissue (a second set of ductoscopy washing samples was not available), we have confirmed that there is a significant difference in the methylation signal of DCIS and benign papilloma from normal breast epithelium. These findings may also have relevance to the identification of women at high risk for breast cancer because benign papillomata with a high level of DNA methylation may be particularly prone to malignant progression.

pathology will enable *in situ* ablation of benign lesions with either endoscopic techniques or the intraductal administration of antineoplastic agents (19).

Gene promoter hypermethylation is a hallmark of breast cancer (20). We recently developed an assay that, in the ductal lavage setting, detected breast cancer cells in cytologic fluids with over twice the sensitivity of cytology alone (21). The method is based on the assessment of levels of promoter methylation in genes using a technique called quantitative multiplex-methylation–specific PCR (QM-MSP). The QM-MSP assay can evaluate a panel of up to 12 genes in patient samples containing as few as 500 to 1,000 cells. We proposed that use of this biomarker panel to evaluate ductal cells could help to distinguish between PND ducts with a malignant neoplasm and those with benign lesions. This could provide a rationale to avoid diagnostic ductal excision in most women with PND. Furthermore, it is possible that ducts with benign lesions that display high levels of methylation are at higher risk for developing future breast cancer. These ducts could be targeted for more careful surveillance. To test this concept, we did a prospective study to evaluate whether quantitative assessment of gene promoter hypermethylation could enhance the utility of office ductoscopy in women with PND and those with atypical cytology on ductal lavage (5). In this article, we compare methylation levels of a panel of 11 genes in cells obtained from ductoscopy washings derived from ducts with minimal or no ductoscopy findings versus ducts with clinically significant findings. Among ducts with significant gross lesions, we refine this panel to determine which genes, when assessed cumulatively, enhance the positive predictive value (PPV) of ductoscopy for detection of DCIS in women with PND.

Materials and Methods

Patients. Women presenting to the Bluhm Family Program for Breast Cancer Early Detection and Prevention at Northwestern Memorial Hospital were eligible for participation in the study if they had PND ($n = 45$), or atypical ductal lavage cytology ($n = 4$). We included women who showed two of the three signs that characterize pathologic discharge: spontaneity, emanation from a single duct, and bloody or serous character. All participants underwent ductoscopy. The patients' ages ranged from 26 to 75 y (mean, 48 y). Patients provided informed consent and use of all samples for this study was approved by the Institutional Review Boards at Northwestern University and Johns Hopkins University. The flow of clinical procedures and associated samples is laid out In Fig. 1.

Ductoscopy and collection of cells. Ductoscopy was done as previously described (5). If ductoscopy revealed significant findings (1 or more papillomatous growths), the patient underwent surgical resection (36 ducts in 33 women; mean ducts per woman, 1.1). If the ductoscopy findings were normal, or if insignificant "abnormalities" were present (wispy fronds or flat red patches), clinical follow-up was offered, without surgical resection (28 ducts in 16 women; minimum 2-y follow-up; mean ducts per woman, 1.8). Of the 49 women who underwent ductoscopy, 35 (71%) had 1 duct and 14 (29%) had 2 or 3 ducts examined. Eight of these 14 women (57%) were in the surgical group and 6 (43%) were in the nonsurgical group. In the surgical group, one of the eight women with more than one duct examined showed DCIS in the surgical specimen, whereas seven of eight were found to have a papilloma. Samples from two ducts failed methylation analyses because of insufficient cellularity.

Cytology. When abnormalities were visualized ductoscopically, brushings of these were collected, fixed in CytoLyt, and stored in PreservCyt (Cytyc Corporation). One-tenth of the sample was used to prepare a Papanicolaou-stained slide for cytologic evaluation. The remainder was used for QM-MSP after centrifuging the cells and removing any fluid to determine the incidence and level of methylation. Cytology was classified as benign, mildly atypical, severely atypical, and malignant (21, 22). A positive cytologic test was defined as severely atypical or malignant (21).

DNA extraction. Extraction of DNA was done as described (21, 23) on cells obtained from ductoscopic washings, paraffin-embedded tissue sections, and breast organoids. Fluid DNA from pathologically discharging nipple fluid (1-30 μL) was prepared using the High Pure Viral DNA Method (Roche Applied Bioscience; Any nipple fluid cells suspended in the fluid were included in the prepared DNA). Before DNA isolation, normal breast organoids were prepared by mincing of fresh breast tissue and partial digestion of the tissue overnight at 37°C with gentle inversion in 1.6 mg/mL collagenase A (Roche Applied Science) and 2.5 μg/mL hyaluronidase in DMEM/F12 containing 1% bovine serum albumin, 1 μg/mL insulin, and antibiotics. Sodium bisulfite treatment of genomic DNA was done according to the method previously reported (21).

Validation set. The ductoscopy washing test set was developed with pilot study funding from the Early Detection Research Network of the National Cancer Institute. Because office ductoscopy is not generally reimbursed by third party payors, and DCIS presenting as nipple discharge is a relatively rare event, we turned to pathology archives to develop a validation set. Through clinical records, we identified women who presented for surgical duct excision because of pathological nipple discharge, and were found to have papilloma or DCIS in the surgical specimen. They were treated over the same time period as the women in the test set, and had a similar age range 31 to 73 y (mean, 49 y). We processed samples "as is," without further enrichment of lesion tissue to ascertain in routine tissue sections the extent

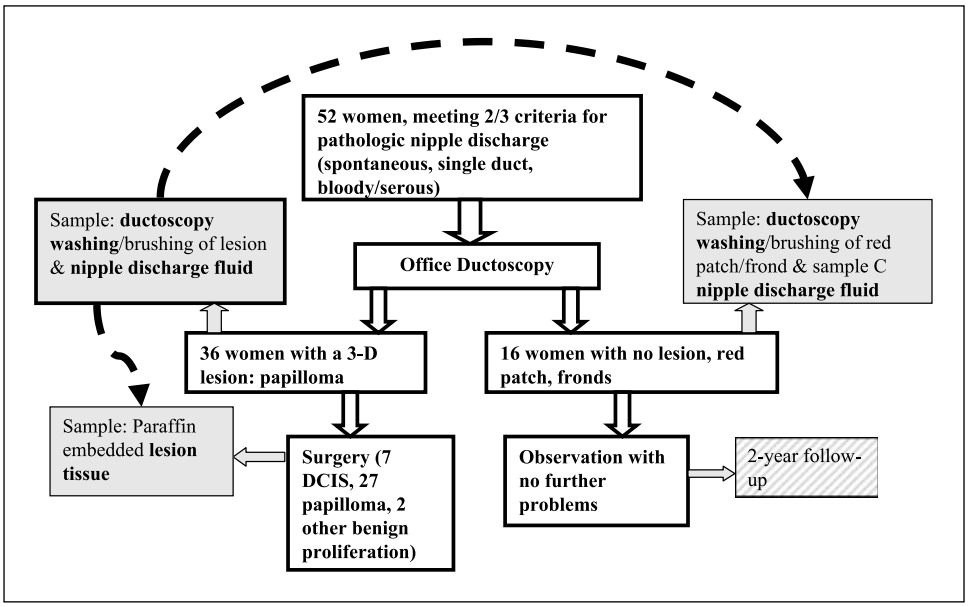

Fig. 1. The flow of patients and samples is illustrated; open arrows, clinical events; filled arrows, samples derived from these; broken arrows, comparisons between groups of patient samples. Forty-nine women underwent office ductoscopy. Of these, 33 had three-dimensional grossly papillomatous lesions and proceeded to surgical resection. Histologic examination of lesions showed 7 DCIS (1 with invasion), 27 papillomas, 1 atypical ductal hyperplasia, and 1 ductal hyperplasia of the usual type. Sixteen women showed completely normal ducts, flat red patches, or wispy fronds with no three-dimensional structure to them, and were observed for a minimum of 2 y. The ductoscopy washings were studied for cytology and QM-MSP, and findings were compared between the surgical and observation groups. The QM-MSP ductoscopy washing data from the 33 women with significant lesions were compared with (*a*) those with normal ducts/insignificant lesions (Figs. 2 and 3), (*b*) with histologic sample QM-MSP (Fig. 4), and (*c*) with nipple discharge QM-MSP (Fig. 4).

of methylation differences between cases of cancer and papilloma. Methylation analyses were done blinded to the case identity, histologic diagnosis or % lesion within the tissue The normal comparison group was derived from 17 women undergoing reduction mammoplasty; parenchymal tissue (ducts and terminal ductal units) were depleted of surrounding stromal tissue to yield samples consisting of 70% to 80% epithelium.

Primers and probes. Sequences are available upon request.

QM-SP methodology. QM-MSP was done in two sequential PCR steps as described by Fackler et al. (23) and Swift-Scanlan et al. (2006): Step 1: Multiplex PCR was done using external primers independent of DNA methylation status to coamplify up to 11 genes. The final PCR product was then diluted 1:5 or more in sterile dH_2O, depending on the concentration of input DNA. Step 2: In

Table 1. Gene promoter hypermethylation

	RASSF1A	HIN1	TWIST1	CYCLIN D2	APC1	RAR β	CDH1	BRCA1	BRCA2	ER α	P16
Cancer											
No. of values	7	7	7	7	7	7	6	6	7	7	7
25% percentile	1.08	0.80	0.45	0.11	0.10	0.06	0.01	0.00	0.00	0.00	0.00
Median	1.84	1.95	5.23	1.21	0.26	2.52	0.17	0.00	0.00	0.00	0.00
75% percentile	35.00	40.32	30.95	37.66	16.09	14.45	4.19	0.83	0.00	0.00	0.33
Mean	13.92	15.69	18.02	15.31	10.23	9.96	2.63	0.40	0.00	0.01	0.99
Papilloma											
No. of values	27	26	27	27	25	27	22	27	27	26	27
25% percentile	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00
Median	0.44	0.02	0.03	0.01	0.00	1.07	0.00	0.00	0.00	0.00	0.00
75% percentile	1.42	0.68	0.11	0.31	0.25	2.07	0.10	0.00	0.00	0.00	0.00
Mean	3.27	1.33	0.61	0.54	1.56	1.46	3.59	2.40	0.01	0.01	0.00
Normal											
No. of values	28	28	28	28	28	28	22	27	28	28	28
25% percentile	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00
Median	0.05	0.00	0.03	0.26	0.01	0.94	0.00	0.00	0.00	0.00	0.00
75% percentile	0.52	0.01	0.53	0.65	0.25	2.05	0.02	0.00	0.00	0.00	0.00
Mean	0.29	0.15	0.35	0.53	0.28	1.40	0.05	0.00	0.05	0.02	0.00
P											
Normal vs cancer	0.001	0.000	0.002	0.061	0.062	0.312	0.020	0.022	0.020	0.397	0.130
Papilloma vs cancer	0.017	0.019	0.001	0.009	0.034	0.394	0.137	0.217	0.137	0.323	0.283
Normal vs papilloma	0.167	0.085	0.946	0.185	0.204	0.866	0.383	0.286	0.383	0.824	0.197

NOTE: Cells were harvested during ductoscopy and the % methylation was determined by QM-MSP. Shown is the median, quartiles, and mean level of % methylation for 11 genes (*P* values calculated by the Mann-Whitney *U* Test).

the quantitative reaction, 2-color real-time MSP was done on 1 μL of the diluted PCR product from step 1. The second step was carried out in a 25-μL PCR final reaction volume with 1.25 units of Platinum DNA Taq Polymerase (Invitrogen), 1× ROX (Invitrogen) passive reference dye, 1× buffer [16.6 mmol/L $(NH_4)2SO_4$, 67.0 mmol/L Tris (pH 8.8), 6.7 mmol/L $MgCl_2$, 10.0 mmol/L β-mercaptoethanol, 0.1% DMSO], 200 μmol/L deoxynucleotide triphosphate, and 800 nmol/L of each primer set. 6-FAM/TAMRA or VIC/TAMRA-conjugated gene-specific oligonucleotide probes were used with U and M primer sets performing reactions in the same well. The real-time assay was done using an ABI Prism 7900HT Sequence Detector (Applied Biosystems). The standard curve, control, and sample DNAs were prepared essentially as described previously, with the modification that U and M gene MSP standards were cloned into plasmids. Percent methylation for each gene was calculated as [M/(U+M)] ×100 using the absolute quantification method based on a standard curve of copy number. To determine the cumulative methylation index (CMI), the sum of %M for all genes tested within a given sample was determined. For example, for 6 genes, 100%M ×6 genes = CMI of 600 maximum possible methylation units.

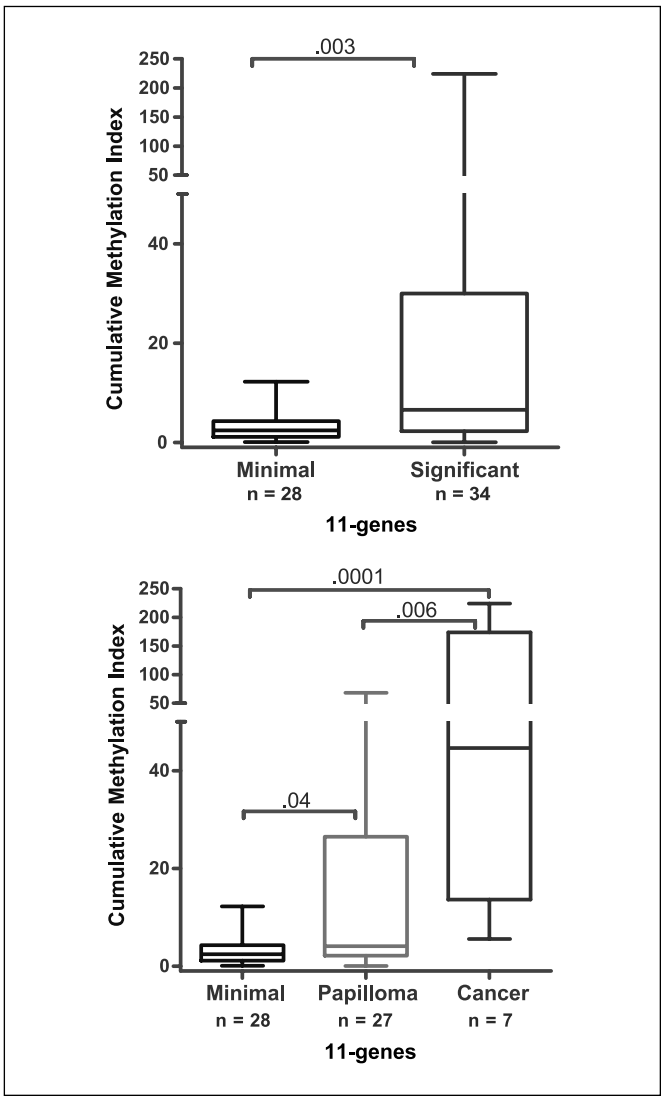

Fig. 2. Gene promoter hypermethylation of *RASSF1A, HIN1, TWIST1, Cyclin D2, APC1, RARβ, CDH1, BRCA1, BRCA2, ERα*, and *p16* in cells harvested during breast ductoscopy. Shown is the level of cumulative methylation within cells washed from ducts with minimal ductoscopic findings (*minimal*) compared with ducts with significant findings (*significant*). *Bottom,* lesions within ducts with significant findings were surgically excised to classify them histologically; significantly different methylation levels were observed in cells washed from ducts with papilloma, cancer, and/or ducts with minimal ductoscopic findings. The median, 25th and 75th quartiles, and range of cumulative methylation are displayed; analyses were done using the Mann-Whitney test.

Statistical analysis. Throughout the analyses, the unit of analysis was the duct. A laboratory normal threshold was defined using the 90th percentile rank method (21) based on the methylation level in ducts with minimal ductoscopy findings. To evaluate the ability of QM-MSP test in further distinguishing between benign and cancerous ducts, a receiver operating characteristic (ROC) approach was used for ducts with significant ductoscopy findings that had definitive diagnosis via tissue histology based on surgery. An optimal threshold was chosen by the highest sensitivity and then maximizing the specificity. CMI was calculated to characterize the methylation profiles for multigene panels. A positive panel was defined as one that exceeded the defined threshold. The operating characteristics parameters of QM-MSP were estimated along with 95% confidence intervals (95% CI). Nonparametric Mann-Whitney *U* tests were used to compare the difference in methylation levels between groups. The analyses of this pilot study are exploratory and results will be validated in an independent, larger cohort of women with PND. Tests were considered to be statistically significant at P value of <0.05, and analyses were done using SAS (version 9.1), and GraphPad (version 5) software packages.

Results

Ducts with minimal or no abnormal ductoscopic findings. Twenty-eight ducts were classified as "normal." These included 7 where lesions were not detectable on ductoscopy, and 21 with visible findings, which were not of clinical concern (8 with scattered flat red patches, and 13 with occasional thin wispy fronds with no bulk to them). QM-MSP was done on the cells obtained from the ductoscopy washings. The percentage of hypermethylated alleles in the normal samples was quite low for all 11 genes (Table 1). All of the seven women who were included in the study because of mild cytologic atypia on ductal lavage had normal or minimal findings on ductoscopy.

Methylation in ducts with significant ductoscopy findings. In 36 ducts from 33 women, significant lesions were detected by ductoscopy. These consisted of solitary or multiple papillomatous lesions with a three-dimensional structure that occupied at least a quarter of the luminal area. Surgery was done to excise and characterize these lesions. Based on histologic evaluation of the excised tissue ("gold standard"), these ducts were then classified into 2 main groups: (*a*) ducts containing malignancy (6 DCIS and 1 DCIS/invasive ductal carcinoma), or (*b*) ducts containing benign lesions (27 papillomas, 1 atypical ductal hyperplasia, and 1 ductal hyperplasia of the usual type). Multiple papillomas were present in 8 of 27 ducts with papilloma. Based on evaluation of 11 candidate genes, the cumulative level of DNA methylation in cells recovered from ducts with significant lesions was higher than in normal ducts ($P = 0.003$; Fig. 2). Furthermore, methylation was significantly higher in samples from ducts proven to contain malignant lesions on resection compared with those with papilloma ($P = 0.006$) or normal appearance ($P = 0.0001$; Fig. 2).

Detection of cancer in ductoscopy washings with QM-MSP. Past studies have shown that a panel of hypermethylated genes was more powerful than single genes in detecting cancer (21, 24, 25). To devise the optimal cancer detection panel, the 6 genes with the best discriminatory ability (*RASSF1A, HIN1, TWIST1, CYCLIN D2, APC1*, and *RARβ*) were identified from among the 11 candidate genes, based on predictive powers

quantified by areas under the ROC curves for normal versus cancer (data not shown). To define our laboratory normal values in washings from ducts with minimal ductoscopy findings, we used the 90th percentile rank of cumulative methylation by QM-MSP (Table 2A; Fig. 3A and B). For the 6-gene panel, the cutoff of 6.2 (of a possible 600 methylation units) yielded a sensitivity of 86% (6 of 7 cancers tested positive) and specificity of 93% (26 of 28 normal ducts tested negative) for an overall classification accuracy of 91%, based on analysis of normal versus cancer (Table 2A). By this analysis, 67% (18 of 27) of the ducts containing papilloma tested negative (within limits of normal). A closer look at the frequency distribution of methylated papilloma samples showed that the highest four samples were outliers (68, 32, 29, and 26 units of methylation, respectively; data not shown).

However, a major diagnostic consideration is to distinguish benign lesions from cancerous lesions within ducts having clinically significant ductoscopy findings. To determine whether QM-MSP analyses could improve performance of the ductoscopy procedure, ROC threshold analyses were done on samples from ducts with significant findings only, comparing methylation in ductoscopy washings from benign ducts versus cancerous ducts (Table 2B). With the 6-gene panel, the threshold was 5.5 units, which yielded a sensitivity of 100% (7 of 7 cancerous ducts), specificity of 66% (19 of 29 benign ducts), classification accuracy of 72% (26 of 36 ductoscopically abnormal ducts), and area under the curve (AUC) of 0.88 (Table 2B).

Selecting a minimal cancer marker panel. Because simpler is better, and to render the assay quicker and more cost effective, we devised a gene panel containing the most informative of the six genes. A 3-gene panel consisting of *RASSF1A, HIN1*, and *TWIST1* genes was selected from the larger panel because they had the highest single-gene predictive power (and AUC of 0.87-0.92) comparing normal ducts to cancerous ducts (Fig. 3B). Comparison of normal versus DCIS ducts, with a laboratory normal cutoff of 2.7 with the 3-gene panel, yielded a sensitivity of 86% (6 of 7 ducts with cancer tested positive), specificity of 93% (26 of 28 normal ducts tested negative), and classification accuracy of 91% (32 of 35 ducts correctly classified; cutoff based on the 90th percentile rank of methylation; Table 2A; Fig. 3B). Seventy-eight percent (21 of 27) of the ducts containing papilloma tested negative.

Among ducts with clinically significant ductoscopy findings, comparing benign lesions to malignant lesions by the ROC threshold method (Table 2B), a 3-gene threshold of 2.6 units yielded a sensitivity of 100% (7 of 7 cancerous ducts detected as positive), specificity of 72% (21 of 29 normal ducts detected as negative), classification accuracy of 78% (28 of 36 ductoscopically abnormal ducts classified correctly), and AUC of 0.91. Based on these data, the PPV of the ductoscopy procedure alone was 19% (7 cancers in 36 ductoscopy/cytology-positive ducts; 95% CI, 8-36%; this was improved to a PPV of 47% by the addition of QM-MSP (7 cancers in 15 QM-MSP–positive ducts; 95% CI, 21-73%). Thus, the PPV of ductoscopy was more than doubled by the addition of QM-MSP. These results suggest that QM-MSP can provide an adjunct to ductoscopy for more accurate selection of the subset of ducts that harbor clinically significant abnormalities for surgical excision.

Cytology versus methylation. Cytologic examination provides an opportunity to detect cancerous cells harvested during ductoscopy. To determine the reliability of cytology, and to compare it to methylation, a three-way comparison was done between lesion histology, ductal cell cytology, and ductal cell methylation. Of the ducts containing histologically confirmed cancer, only 29% (2 of 7; 95% CI, 4-71%) were identified by cytologic exam of the ductal cells (Fig. 3B) and 71% (5 of 7; 95% CI, 29-96%) of ducts with cancer were missed. By contrast, using methylation as a biomarker, 100% (7 of 7; 95% CI, 59-100%; Table 2B; Fig. 3B) of ducts with histologically proven cancer were correctly identified by QM-MSP, i.e., 7 cancers were detected in 36 ductoscopically abnormal ducts. Because cytology detected only 2 of 7 cancerous ducts, sensitivity was 29%. By

Table 2.

A. Normal vs cancer ducts–QM-MSP/90th percentile rank method

Gene panel	Sensitivity (%, ratio; 95% CI)	Specificity (%, ratio; 95% CI)	Classification accuracy (%, ratio; 95% CI)	Papillomas (% neg; 95% CI)	Threshold*
3-gene	86 (6/7; 42-100)	93 (26/28; 77-99)	91 (32/35; 77-98)	78 (21/27; 58-91)	2.7
6-gene	86 (6/7; 42-100)	93 (26/28; 77-99)	91 (32/35; 77-98)	67 (18/27; 46-83)	6.2
11-gene	86 (6/7; 42-100)	93 (26/28; 77-99)	91 (32/35; 77-98)	67 (18/27; 46-83)	7.8

B. Benign vs Cancer ducts -QM-MSP/ROC analysis method

Gene panel	Sensitivity (%, ratio; 95% CI)	Specificity (%, ratio; 95% CI)	Classification accuracy (%, ratio; 95% CI)	AUC (95% CI)	Threshold†
3-gene	100 (7/7; 59-100)	72 (21/29; 53-87)	78 (28/36; 61-90)	0.91 (0.80-1.00)	2.6
6-gene	100 (7/7; 59-100)	66 (19/29; 46-82)	72 (26/36; 55-86)	0.88 (0.76-1.00)	5.5
11-gene	100 (7/7; 59-100)	55 (16/29; 36-74)	64 (23/36; 46-79)	0.84 (0.69-0.99)	5.0

NOTE: 3-gene: RASSF1A, TWIST1, and HIN1; 6-gene: RASSFIA, TWIST1, HIN1, APC1, CYCLIN D2, and RARβ; 11-gene panel: RASSFIA, TWIST1, HIN1, APC1, CYCLIN D2, RARβ, CDH1, BRCA1, BRCA2, and ERα; sensitivity, number of true positives divided by the number of true positives plus false negatives; specificity, number of true negatives divided by the number of true negatives plus false positives; classification accuracy, number of ducts correctly identified/total number of ducts; AUC, the predictive power of the test.

*The upper threshold of reference range of methylation in ductoscopy cells was defined by the 90th percentile rank of methylation in ducts with.

†The upper threshold of benign % methylation in ductoscopy cells was defined by ROC analysis of methylation in ducts.

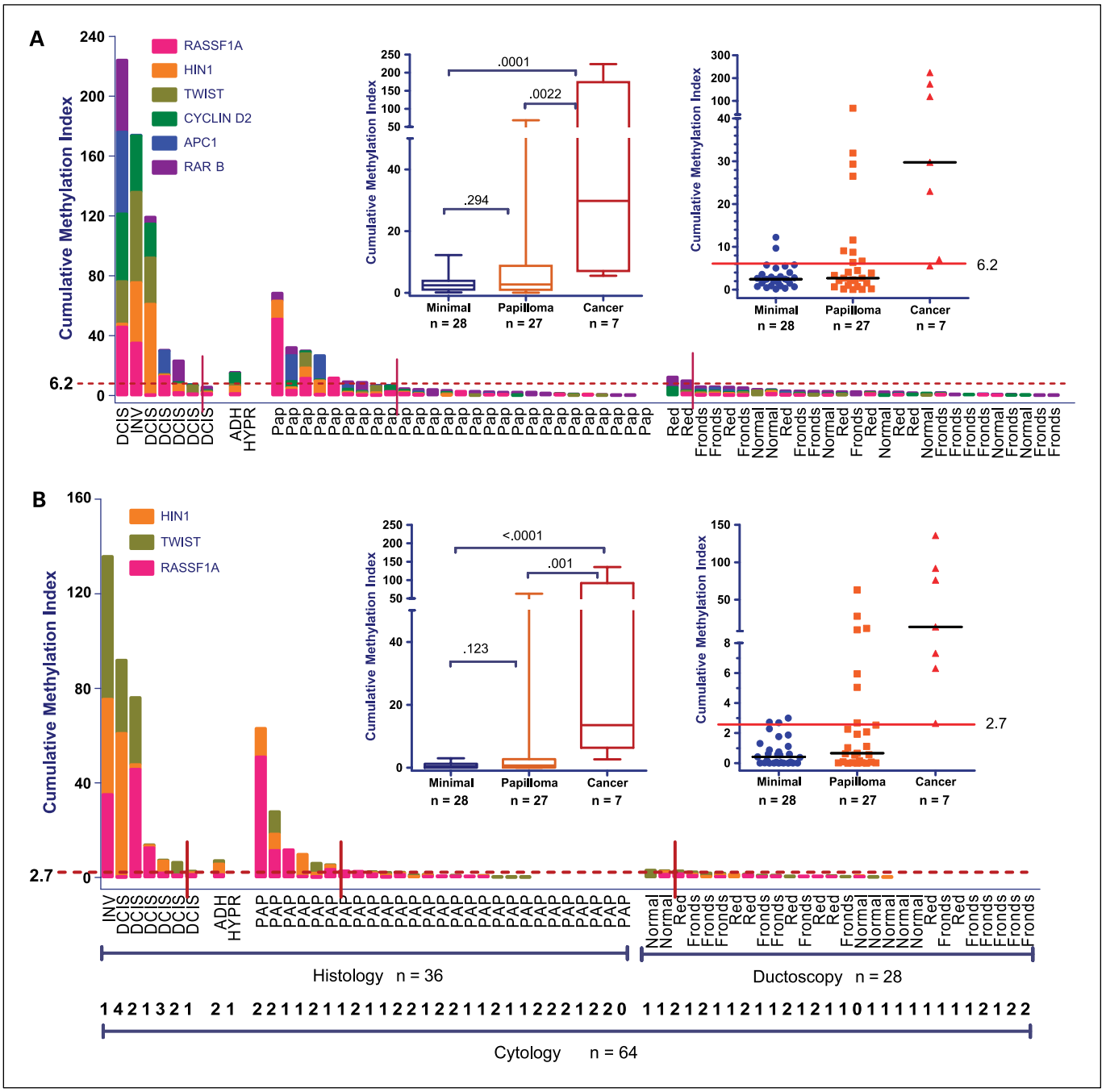

Fig. 3. Detection of cancerous lesions in ducts with PND. The methylation profile of cells washed from individual ducts is shown. The bar height reflects total cumulative methylation (sum of percent methylation of all genes), whereas the color-coded segments correspond to the percent methylation of each gene, as indicated. Dashed line, upper limit of normal methylation, defined by the 90th percentile rank of cumulative methylation in ducts proven to have minimal ductoscopic findings; vertical lines crossing the dashed lines, the cutoff between positive and negative methylation within the group. Histologic classifications were as follows: DCIS, invasive ductal carcinoma (*INV*), atypical hyperplasia (*ADH*), usual hyperplasia (*HYPR*), or papilloma (*Pap*). Ductoscopic classifications (for ducts with minimal findings) were as follows: no lesion (*normal*), reddish lesion (*red*), or wispy fronds (*fronds*). *Insert,* the ability of the gene panel to distinguish between ducts with cancer, papilloma, or normal ducts; The scatter plots indicate the level of cumulative methylation within each sample, the median methylation within the group is shown by the bar in each column and the laboratory threshold (based on the 90th percentile (is indicated by the bar across all three columns in *A* and *B*. *A,* the six-gene detection panel: *RASSF1A, HIN1, TWIST1, CYCLIN D2, APC1,* and *RARβ.* The normal laboratory cutoff value was 6.2 units. *B,* the three-gene detection panel of *RASSF1A, TWIST1,* and *HIN1.* The normal laboratory cutoff was set at 2.7 units. Cell cytology was coded as follows: 4, malignant; 3, severe atypia; 2, mild atypia; 1, benign; 0, insufficient cellular material for diagnosis. Positive cytology was defined as malignant or severe atypia; positive methylation is >2.7 units.

comparison, when the 3-gene QM-MSP test with the ductoscopic washings was added ancillary to the ductoscopy test, the PPV was 47% and sensitivity was 100%. These results using ductoscopic washings support our previous findings on ductal lavage (21), where QM-MSP more than doubled detection of ducts with cancer compared with cytology alone.

Consistency between ductoscopy washings and excised tissue. Ideally, any marker retrieved from cells in ductoscopy

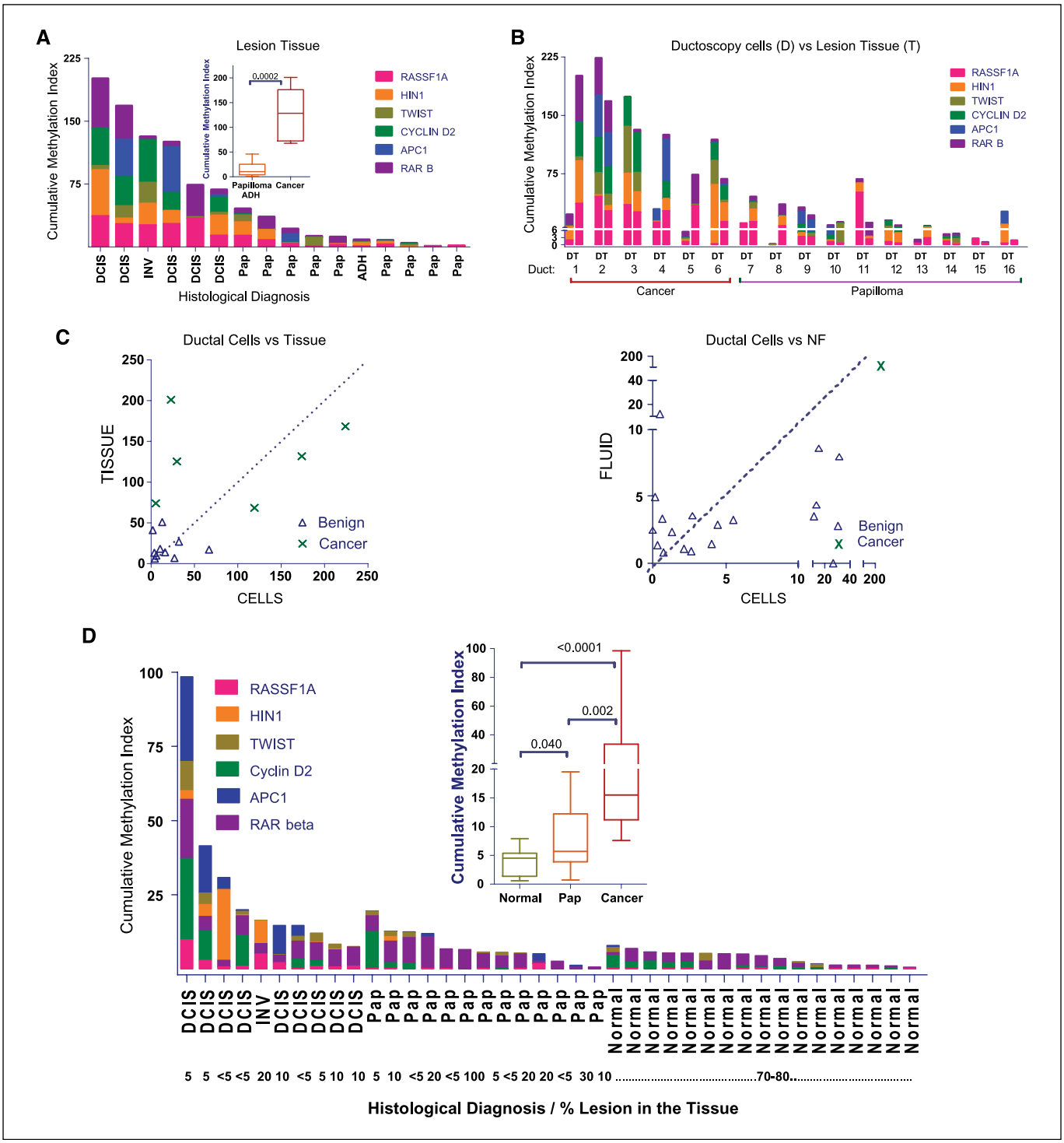

Fig. 4. Methylation in matched tissue and nipple fluids. *A,* in patients with PND ($n = 16$), tissues from excised cancerous and papillomatous lesions were processed for QM-MSP. Results of the methylation profiles of each sample are plotted according to histologic diagnosis (described in Fig. 3). The median methylation in cancerous lesions was significantly higher than for papilloma/atypical hyperplasia (*insert*; $P = 0.0002$). *B,* comparison between tissues (*T*) and harvested ductoscopy cells (*D*). The methylation profile of tissues from *A* was compared with cells harvested at ductoscopy from the same duct (*ducts 1-6,* cancerous; *ducts 7-16,* papillomatous/atypical hyperplasia). *C,* scatter plots by duct: harvested ductoscopic washings (*cells*) versus tissue or nipple fluid. Methylation levels in nipple fluid (*NF*) were substantially lower compared with cells. Dotted line, perfect consistency. *D,* validation of methylation in lesions of patients with PND. In an independent set, patients with PND ($n = 23$) underwent surgical excision of lesions detected by ductoscopy. Excised cancerous ($n = 10$) and papillomatous ($n = 13$) lesions were processed for QM-MSP without further enrichment of lesion tissue to ascertain whether in routine tissue sections, cancer could be distinguished from papilloma by using QM-MSP. As in Fig. 3A, results verify that these lesions are distinguishable. The median methylation in cancerous lesions was significantly higher than for papilloma, (*insert*; $P = 0.002$), as well as for normal breast ($P < 0.0001$). Analyses were done in a manner blinded to the extent of lesion present in the sample. Unblinded data indicated that 9 of 10 cancers and 8 of 13 papilloma samples had ≤10% lesion tissue present in the routine tissue section processed for methylation. The percentage of lesion found within the tissue section processed is indicated at the bottom (normal organoids consist of ~70-80% epithelium) below the histologic diagnosis.

washings should reflect the molecular nature of lesion tissue within the duct. However, in ducts with PND, there is a mixture of cell types, including inflammatory, blood, and normal epithelial cells, as well as tumor cells that may be shed from a lesion into the ductal system. Therefore, we wished to investigate whether the profile of methylated genes was similar within paired sets of ductoscopy cells and excised tissues. To accomplish this, paraffin-embedded tissues were processed from 6 of the 7 cancers (tissue of 1 case was not available), and a random subset of 10 benign lesions. As predicted, cancerous lesions had significantly higher methylation levels than benign lesions ($P = 0.0002$; Fig. 4A). Next, methylation levels in tissue versus ductal cell washings of the same patients were compared. For most cancers, the profile of methylated genes was similar from the two sources in the same patient (example, duct 2; Fig. 4B), although there was variability in the extent of unmethylated DNA in tissue preparations as well as cell washings a (Fig. 4C).

Direct analysis of nipple fluid in patients with PND. Ideally, a noninvasive test for PND diagnosis could be developed, which would detect cancer from the discharge fluid in patients with PND. Therefore, for a random subset of ducts, discharge fluid samples were collected before ductoscopy. Methylation levels in discharge fluid were assessed and compared with matched samples of ductal cells from the same patient retrieved during ductoscopy. QM-MSP analysis for a panel of 11 genes was feasible in 19 of 20 discharge fluid samples (Fig. 4C). Gene methylation levels in the fluid were approximately the same or much lower than in the corresponding ductoscopy sample for nearly all samples (Fig. 4C). In the XY scatter plots shown (Fig. 4C), any mark to the right of the dotted line indicates that there was more methylation in the cells than the fluid. Eleven samples are plotted to the right of the dotted line and seven samples to the left, one is essentially on the dotted line. In general, the samples with highest methylation appear plotted to the right, indicating usually, but not always, that methylation is higher in the cells than the fluid. Overall, Spearmen correlation analyses indicate low correlation between paired samples ($r2 = 0.26$; $P = 0.27$, 2 tailed), in agreement with the idea that cell pellets from ductoscopy washings have different methylation levels compared with the corresponding nipple fluid.

Validation set results. We analyzed paraffin sections containing lesion tissue from 10 women with cancer (9 DCIS and 1 invasive carcinoma) and 13 women with papillomata, all of whom had presented with nipple discharge, and compared them to 17 samples from women undergoing reduction mammoplasty (Fig. 4D). The mean CMI for cancer was 26.5, with a median of 15.5; the mean CMI for papillomata was 7.4, with a median of 5.7, of 600 units possible. These were significantly different ($P = 0.002$). When papilloma samples were compared with reduction mammoplasty samples, the CMI of papilloma was significantly higher ($P = 0.04$), whereas the CMI between reduction mammoplasty samples and cancer was strikingly different ($P < 0.0001$; Mann-Whitney test). By design, we processed samples "as is," without further enrichment of lesion tissue to ascertain whether in routine tissue sections cases of cancer could be distinguished from papilloma by QM-MSP. On pathology review, 9 of 10 cancers and 8 of 13 papilloma samples had ≤10% lesion tissue present in the routine tissue section processed for methylation (Fig. 4D). Despite the under-representation of lesion amidst normal unmethylated stroma/epithelium it was possible to distinguish between cancer, papilloma, and normal breast with high significance using QM-MSP.

Discussion

Currently, no existing test reliably distinguishes between DCIS and benign changes in women with nipple discharge, and the standard of care at this time is to completely excise all papillomas because they can be heterogeneous and associated with malignancy. We investigated the hypothesis that quantitative assessment of hypermethylated genes by QM-MSP can enhance ductoscopy detection of cancer in ducts of women with PND. Our results show that QM-MSP readily distinguished normal ducts, those with minimal ductoscopy findings, from those with significant findings. More importantly, in ducts with significant intraductal lesions, QM-MSP distinguished ducts harboring benign lesions from those containing cancer. Based on the results of this study, QM-MSP may aid in selection of patients for surgical resection, and predict the presence of DCIS rather than papilloma in those undergoing surgery. In addition, as endoscopic ablation tools become available, the ability to distinguish malignancy from benign change will be of crucial importance because *in situ* ablation will logically be more applicable to benign lesions.

The clinical criteria for surgical duct excision are based on the presence of spontaneous single duct discharge that is either bloody or serous. Using these criteria, the yield of pathologic findings in excised ducts is highly variable. We designed our study to include women with discharge that fulfilled two of these three criteria to reexamine the threshold for which excision should be recommended, and to broaden the study population to allow analyses of sensitivity and specificity. As a result, we encountered ducts without dominant or significant intraluminal lesions, which fulfilled at least two of the three criteria for PND. The mean number of ducts examined in the nonsurgical group was higher than the mean number of ducts examined in the surgical group (1.8 versus 1.1); this is consistent with the fact that single duct discharge is a feature of PND and therefore women without significant findings would be more likely to present with more than one discharging duct. In samples that were retrieved from the ductoscopically "normal" ducts in the nonsurgical group, levels of gene methylation were very low for all 11 genes (Table 1); this level was significantly lower than ducts with lesions ($P = 0.003$; Fig. 2).

When compared with normal ducts, the percent methylation was significantly higher in cancerous ducts ($P = 0.0001$), or ducts with papilloma ($P = 0.04$) than in normal ducts (Fig. 2). Most importantly, significant differences were also seen between the methylation levels in cancerous ducts and benign papilloma ($P = 0.006$). It is important to emphasize that QM-MSP output is a continuous variable (percent methylation). It is therefore possible to evaluate the level and incidence of hypermethylation of a single gene, as well as that of a gene panel. Furthermore, based on a reference set of samples, in this pilot study, we established our normal laboratory range and upper cutoff for methylation in PND ductoscopy samples. For the purposes of the present PND study, the normal cutoff was higher than in our previous findings in ductal lavage of normal ducts (albeit in high risk

women; ref. 21), probably because here the cutoff was based on clinically symptomatic ducts. Although this cutoff applies only to studies done within our laboratory, a similar approach can be easily adapted by other investigators and clinical laboratories (26). However, this threshold difference suggests that although cumulative DNA methylation of several genes is a powerful biomarker, the specific genes and the specific threshold may vary depending on the type of sample.

Our previous studies clearly shown that a gene panel has a higher predictive power than a single gene when detecting breast cancer (21). In the current study of cells retrieved during ductoscopy, 11 genes were analyzed, and through an iterative procedure, we identified a panel of the three best performing genes that distinguished between cancerous and ductoscopically normal ducts [*RASSF1A* (AUC, 0.92; $P = 0.0006$); Table 2], *TWIST1* (AUC, 0.89; $P = 0.0017$), and *HIN1* (AUC, 0.87; $P = 0.0026$; data not shown). As a panel, this group of genes was highly accurate at distinguishing ducts without significant lesions on ductoscopy from ducts with cancer (91% classification accuracy; Table 2A). Using the laboratory normal cutoff of 2.7 units, 78% (21 of 27) of papilloma tested negative and 86% (6 of 7) of cancers tested positive.

Although most intraductal lesions can be successfully visualized with ductoscopy, the real challenge lies in the distinction between benign lesions and malignant lesions (Table 2B). To evaluate the QM-MSP test in aiding this distinction, we did an ROC analysis that enabled us to define the optimal threshold of 2.6 CMI units for desirable test performance. With this approach, 3-gene panel detected all of the cancers (100% sensitivity) and 72% of papillomas tested negative. Thus, the combination of ductoscopy plus QM-MSP provides a high classification accuracy (78%); 28 of 36 ducts with significant ductoscopic lesions were correctly classified, with 72% specificity (Table 2B). This high predictive power (AUC, 0.91) and improved PPV (47%), suggests an incremental value of QM-MSP to ductoscopy for detection of cancerous lesions in women with PND. Of interest, the 3-gene panel had higher specificity than the 6- or 11- gene panels while maintaining a 100% sensitivity, demonstrating that the use of fewer genes can produce a more specific marker panel, reducing the time and cost involved in diagnosis.

Among women presenting with nipple discharge, cytologic examination has low sensitivity and specificity and is therefore variably useful (4, 27). It also requires access to a highly trained cytopathologist. Furthermore, we found in this and previous studies of ductal lavage (21) that QM-MSP detected cancer more often than cytology. In the present study, less than one-third (two of seven) of cancer-containing ducts were identified by cytologic examination of ductoscopy washings (Fig. 3B). By contrast, QM-MSP identified 100% of cancerous ducts (7 of 7), thus performing with more than twice higher sensitivity than cytology, as we previously reported (21). Furthermore, a few papilloma samples, although negative by cytology, showed levels of methylation that were as high as those observed in ducts with cancer (Fig. 3). This raises the question of whether, left alone, these are the papillomas that may in the future give rise to DCIS and the observed high methylation is a reflection of molecular changes associated with cancer, which precede changes in cell morphology. These findings lead to the hypothesis that methylation changes quantitated by QM-MSP have the potential to serve as indicators of cancer risk.

Having observed that methylation is significantly higher in samples from ducts with DCIS, we next sought to show that these findings are representative of the abnormal lesion in the duct by performing QM-MSP analyses on tissue samples from matched, excised lesions. We showed that tissue from cancerous breast lesions had significantly higher methylation than from papillomata ($P = 0.0002$; Fig. 4A). Upon comparison of matched DNA from ductoscopy cells and tissue sections (Fig. 4B and C) in cancerous ducts, essentially the same genes were seen hypermethylated in both duct washings and the ductal lesion. This finding strongly suggests that in ducts containing cancer, the tumor cells are shed into the duct from the lesion and are detected by QM-MSP.

The three-gene QM-MSP test would be more practical if it could be applied directly to the nipple discharge fluid in patients with PND. Our results showed that the assay was successful in most discharge samples, but that in many of these samples methylation values in the fluid were less robust than in the ductoscopy cell sample (Fig. 4C). This is not surprising because the cellular content of nipple discharge is much lower than in ductoscopy washings; additionally, the act of brushing the lesion during ductoscopy may result in an abundance of hypermethylated DNA to provide a suitable template for sensitive detection of cancer. Our results also likely indicate that the majority of methylation signal comes from the atypical cells in those samples and the majority of DNA from the nipple fluid may come from apoptotic or necrosed inflammatory cells. However, some highly cellular nipple fluid samples may be the exception. The goal of lesion detection and diagnosis using nipple discharge samples remains important and awaits further technological refinements.

Validation of these results in an independent sample set is crucial, but additional ductoscopy washings were not available to us because the funding for performance of office ductoscopy was limited to a pilot study. We chose therefore to evaluate the QM-MSP levels in archival paraffin-embedded tissue samples from women who had undergone duct excision for PND. These paraffin embedded samples (Fig. 4D) were therefore from a population identical to the tissues in the test set (Fig. 4A), and both sets show striking and highly significant differences in methylation between benign proliferation (papilloma/atypical hyperplasia) and cancer. We felt that this second set of tissue from the same clinical population would provide initial validation of our findings that benign lesions and papilloma are distinguishable by methylation. By design, in our validation set, we processed samples without further enrichment of lesion tissue to ascertain whether in routine tissue sections cancer could be distinguished from papilloma by QM-MSP (Fig. 4). Despite the under-representation of lesion amidst normal unmethylated stroma/epithelium in these samples that had not undergone microdissection, it was possible to distinguish between cancer, papilloma, and normal breast with high significance using QM-MSP ($P < 0.0001$). However, in future studies to ensure uniformity between samples, we plan to enrich for the lesion tissue by scraping away grossly normal tissue in the tissue section. QM-MSP will be especially important for smaller samples of tissue, such as are obtained through fine needle biopsy, where histologic diagnosis is challenging and an ancillary test for cancer would be useful.

In conclusion, we have shown that the measurement of DNA methylation in cell samples obtained using ductoscopy in women with PND is feasible; and that there is a significant gradient in the levels of DNA methylation in samples from ducts

with malignancy, those with benign lesions, and those with no significant findings on ductoscopy. These results are promising in terms of the potential of QM-MSP measurement of DNA methylation as a diagnostic test in women with nipple discharge; and may allow stratification of women with benign papillomata into those at high or low risk for future breast cancer based on DNA methylation levels. Prospective larger validation studies are being planned in women with PND to define the role of this powerful tool in the early detection and risk assessment of breast cancer. Furthermore, these results show proof of the principal that QM-MSP performs well in the challenging distinction of benign proliferation from *in situ* carcinoma, a very challenging threshold in the development of molecular diagnostic tests.

Disclosure of Potential Conflicts of Interest

The QM-MSP method is licensed by OncoMethylome Sciences. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.

Acknowledgments

We thank Nancy Davidson, Antonio Wolff, and Vered Stearns for their thoughtful reviews of this paper.

References

1. Dixon JM, Mansel RE. ABC of breast diseases. Symptoms assessment and guidelines for referral. BMJ 1994;309:722–6.
2. Gulay H, Bora S, Kilicturgay S, Hamaloglu E, Goksel HA. Management of nipple discharge. J Am Coll Surg 1994;178:471–4.
3. Dawes LG, Bowen C, Venta LA, Morrow M. Ductography for nipple discharge: no replacement for ductal excision. Surgery 1998;124:685–91.
4. Cabioglu N, Hunt KK, Singletary SE, et al. Surgical decision making and factors determining a diagnosis of breast carcinoma in women presenting with nipple discharge. J Am Coll Surg 2003;196:354–64.
5. Moncrief RM, Nayar R, Diaz LK, Staradub VL, Morrow M, Khan SA. A comparison of ductoscopy-guided and conventional surgical excision in women with pathologic nipple discharge. Ann Surg 2005;241:575–81.
6. Liu GY, Lu JS, Shen KW, et al. Fiberoptic ductoscopy combined with cytology testing in the patients of pathologic nipple discharge. Breast Cancer Res Treat 2007.
7. Shen KW, Wu J, Lu JS, et al. Fiberoptic ductoscopy for patients with nipple discharge. Cancer 2000;89:1512–9.
8. Gray RJ, Pockaj BA, Karstaedt PJ. Navigating murky waters: a modern treatment algorithm for nipple discharge. Am J Surg 2007;194:850–4, discussion 4-5.
9. Sauter ER, Ehya H, Klein-Szanto AJ, Wagner-Mann C, MacGibbon B. Fiberoptic ductoscopy findings in women with and without pathologic nipple discharge. Cancer 2005;103:914–21.
10. Morrogh M, Morris EA, Liberman L, Borgen PI, King TA. The predictive value of ductography and magnetic resonance imaging in the management of nipple discharge. Ann Surg Oncol 2007;14:3369–77.
11. Dua RS, Isacke CM, Gui GP. The intraductal approach to breast cancer biomarker discovery. J Clin Oncol 2006;24:1209–16.
12. Sauter ER, Ehya H, Schlatter L, MacGibbon B. Ductoscopic cytology to detect breast cancer. Cancer J 2004;10:33–41, discussion 15-6.
13. Dooley WC, Francescatti D, Clark L, Webber G. Office-based breast ductoscopy for diagnosis. Am J Surg 2004;188:415–8.
14. Hunerbein M, Raubach M, Gebauer B, Schneider W, Schlag PM. Ductoscopy and intraductal vacuum assisted biopsy in women with pathologic nipple discharge. Breast Cancer Res Treat 2006;99:301–7.
15. Al Sarakbi W, Salhab M, Mokbel K. Does mammary ductoscopy have a role in clinical practice?. Int Semin Surg Oncol 2006;3:16.
16. Danforth DN, Jr., Abati A, Filie A, et al. Combined breast ductal lavage and ductal endoscopy for the evaluation of the high-risk breast: a feasibility study. J Surg Oncol 2006;94:555–64.
17. Valdes EK, Boolbol SK, Cohen JM, Balassanian R, Feldman SM. Clinical experience with mammary ductoscopy. Ann Surg Oncol 2006.
18. Okazaki A, Okazaki M, Asaishi K, et al. Fiberoptic ductoscopy of the breast: a new diagnostic procedure for nipple discharge. Jpn J Clin Oncol 1991;21:188–93.
19. Murata S, Kominsky SL, Vali M, et al. Ductal access for prevention and therapy of mammary tumors. Cancer Res 2006;66:638–45.
20. Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. Oncogene 2002;21:5462–82.
21. Fackler MJ, Malone K, Zhang Z, et al. Quantitative multiplex methylation-specific PCR analysis doubles detection of tumor cells in breast ductal fluid. Clin Cancer Res 2006;12:3306–10.
22. Khan SA, Wiley EL, Rodriguez N, et al. Ductal lavage findings in women with known breast cancer undergoing mastectomy. J Natl Cancer Inst 2004;96:1510–7.
23. Fackler MJ, McVeigh M, Mehrotra J, et al. Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. Cancer Res 2004;64:4442–52.
24. Evron E, Dooley WC, Umbricht CB, et al. Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. Lancet 2001;357:1335–6.
25. Locke I, Kote-Jarai Z, Fackler MJ, et al. Gene promoter hypermethylation in ductal lavage fluid from healthy BRCA gene mutation carriers and mutation-negative controls. Breast Cancer Res 2007;9:R20.
26. Euhus DM, Bu D, Ashfaq R, et al. Atypia and DNA methylation in nipple duct lavage in relation to predicted breast cancer risk. Cancer Epidemiol Biomarkers Prev 2007;16:1812–21.
27. Markopoulos C, Mantas D, Kouskos E, et al. Surgical management of nipple discharge. Eur J Gynaecol Oncol 2006;27:275–8.

Clinical Cancer Research

Hypermethylated Genes as Biomarkers of Cancer in Women with Pathologic Nipple Discharge

Mary Jo Fackler, Aeisha Rivers, Wei Wen Teo, et al.

Clin Cancer Res 2009;15:3802-3811.