AD_____

Award Number: DAMD17-00-1-0132

TITLE: Protein Microarray Technology for the Noninvasive Diagnosis and Prognosis of Breast Cancer

PRINCIPAL INVESTIGATOR: Richard C. Zangar, Ph.D.

CONTRACTING ORGANIZATION: Battelle
Richland, Washington 99352

REPORT DATE: July 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2003	Final (1 Jul 2000 - 30 Jun 2003)

4. TITLE AND SUBTITLE
Protein Microarray Technology for the Noninvasive Diagnosis and Prognosis of Breast Cancer

5. FUNDING NUMBERS
DAMD17-00-1-0132

6. AUTHOR(S)
Richard C. Zangar, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Battelle
Richland, Washington 99352

E-Mail: richard.zangar@pnL.gov

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

A number of circulation markers have been identified that have the potential to be used in the detection or prognosis of breast cancer. Unfortunately, no single marker is consistently increased in breast cancer patients when compared with the general population. We hypothesized, however, that a sophisticated analysis of large number of circulation markers would accurately detect breast cancer as well as provide a valuable tool for prognosis. Therefore, we proposed to develop a rapid and simple system to measure a large number of blood markers associated with breast cancer. In order to accomplish this we have developed an antibody microarray with antibodies specific to different blood markers. We have sceened twenty markers and generated. We have refined the microarry to measure markers with a sensitivity down to 0.5 pg/ml. We have used employed this microarry to 200 serum samples form breast cancer patients and control patients. These data have undergone an initial analysis and a number of relationships have been identified. These data will be analyzed using sophisticated computer programs that are designed to find relationships in a complex data set such as this. These studies will result in a prototype chip that can be used for the rapid determination of circulation markers associated with breast cancer.

14. SUBJECT TERMS
Detection, diagnosis and prognosis, monoclonal antibody microarray chip, circulating markers, bioinformatics

15. NUMBER OF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Circulating blood carries chemical information from every cell in the body in the form of proteins, hormones and other factors that can potentially be assayed to screen for cancers and other diseases. In the case of breast cancer, a number of circulating markers have been identified that have the potential to be used in the detection or prognosis of the disease. Unfortunately, no single marker is consistently increased in breast cancer patients when compared with the general population. We hypothesized, however, that a sophisticated analysis of large numbers of circulating markers would accurately detect breast cancer as well as provide a valuable tool for prognosis. Therefore, we proposed to develop a rapid and simple system for this purpose. We have accomplished this by developing an antibody microarray with antibodies specific to nineteen different markers. We have refined the microarray to measure markers with a sensitivity down to 0.5 pg/ml which is comparable to a good commercial 96-well ELISA. We have employed this microarry to screen 200 serum samples from breast cancer patients and control patients. These data have undergone an initial analysis and a number of relationships have been identified. These studies have resulted in a prototype chip that can be used for the rapid determination of circulating markers associated with breast cancer. This basic technology is likely to lead to the development of more advanced chips with wide application in screening, diagnosis, and prognosis of patients with breast cancer.

Body

We made significant progress toward accomplishing the tasks outlined in our statement of work during this project. Task #1 (reprinted from our approved Statement of Work). "Design and test a diagnostic protein chip containing a repertoire (up to 25) of monoclonal antibodies specific to serum tumor markers associated with breast cancer (months 1-24)."

- Develop a microarray chip containing up to 25 different antibodies that recognize circulating markers associated with breast cancer. *Completed*
- Collect a preliminary number of serum sample from individuals that are apparently cancer-free and those with breast cancer. We estimate that we will have about 30-50 samples of each type by this time. These samples will be screened by Western blot

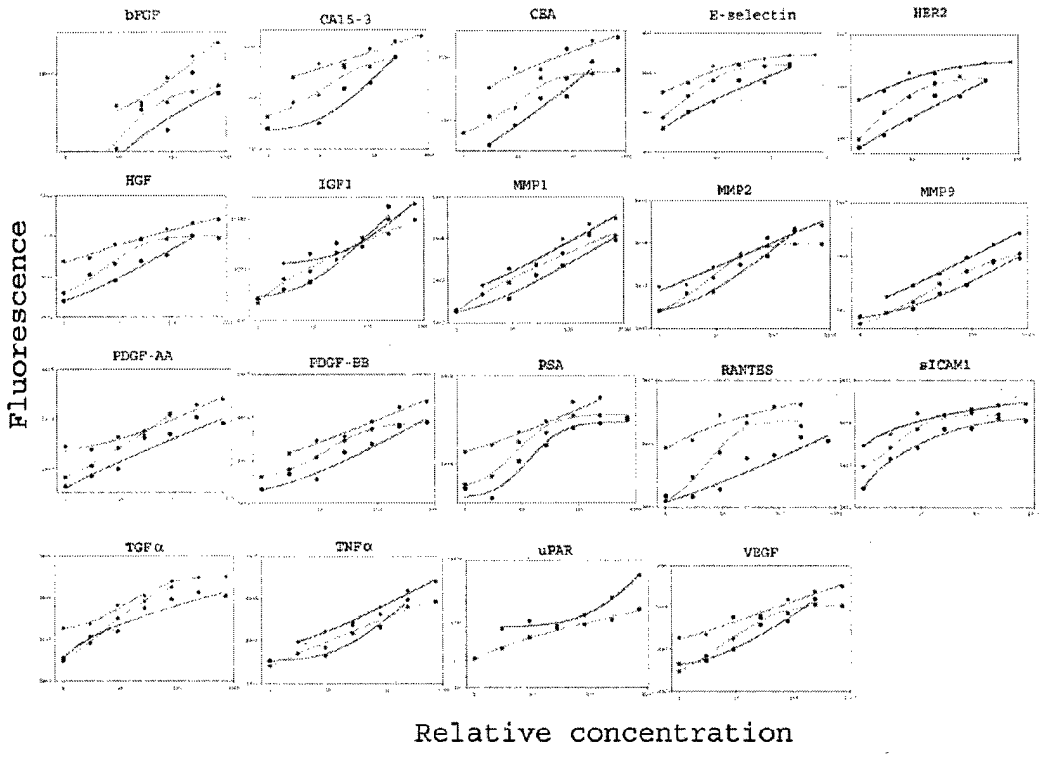
methods to identify samples which have high and low levels of each targeted marker. *Completed*

- Test the microarray chip using the sera identified in the above step. This will allow us to determine appropriate conditions for detection. Factors that potentially may be varied are amounts of antibodies used, either for binding to the spot or for detection; dilution of serum; incubation time; and source of antibody (some antibodies may not work satisfactorily). *Completed*
- Day to day reproducibility and stability of the chips will also be determined. *Partially completed*

We initially refined the microarray format using hepatocyte growth factor (HGF) as a test protein for detection. The microarray can detect HGF at sub-pg/ml concentrations in sample volumes of 100 microliters or less. Additionally, we showed that the microassay is quantitative and used the microassay to detect elevated HGF levels in sera from recurrent breast cancer patients. We also showed that multiple biomarkers can be simultaneously measured on a single microarray. This work was published in the *Journal of Proteome Research* and is included here as Appendix 1.

During the course of this project we have acquired the antibodies and antigens to quantitatively measure the levels of 19 breast cancer biomarkers: CA15-3, carcinoembronic antigen (CEA), E-Selectin, Fas-ligand, fibroblast growth factor (bFGF), HER-2, HGF, I-CAM, MMP1, MMP2, MMP9, platelet derived growth factors-AA and -BB (PDGF-AA, PDGF-BB), prostrate specific antigen (PSA), RANTES, transforming growth factor alpha (TGF-α), tumor necrosis factor (TNFα), μPAR, and vascular endothelial growth factor (VEGF). Standard curves were generated for each marker (Figure 1) and the quantitative range for each marker is between 2 to 3 orders of magnitude with the sensitivity ranging from sub-pg/ml to 10 pg/ml. Furthermore we are able to quantitate these markers within the expected physiological range for each marker. We were unable to get reproducible data from the microarray ELISA for detecting cathepsin D and osteopontin.

Figure 1. Standard curves for nineteen breast cancer biomarkers measured on three different days. Day1 corresponds to red line, day 2 corresponds to green line, and day 3 corresponds to purple line.



In this year we made significant progress in addressing Task #2 (reprinted here from our approved Statement of Work). "Analyze approximately 100 serum samples from breast cancer patients and 100 from apparently healthy individuals for levels of the marker proteins. This data will then be analyzed using conventional statistics and bioinformatics software (SPIRE) developed at this institute to delineate associations between circulating markers and the presence and stage of breast cancer (months 25-36)."

- The 200 serum samples will be analyzed using the microarray mAb chip developed in task 1. *Completed*
- The data will be analyzed using the SPIRE software and conventional statistics. *Partially completed*
- The resulting data will be used to evaluate the usefulness of the chip in the detection and prognosis of breast cancer as well as determining the contribution of individual markers to assessing breast cancer. *Partially completed*

The microarray ELISA was expanded to include antibodies to the 19 breast cancer biomarkers listed above. Using this microarray we screened approximately 200 serum samples; 65 normal controls, 50 samples from high risk woman, 45 samples from woman diagnosed with stage I or stage II breast cancer and 39 samples from woman with recurrent breast cancer (stage III and IV). The quantitative level of each serum biomarker was determined for each patient and the values from patients in the same group (i.e. normal control, high risk, stage I and II, and stage III and IV) were averaged together. The results from this analysis are shown in Table 1.

Table 1. Average percent of each potential breast cancer-specific biomarker with respect to normal risk average. The microarray ELISA was used to measure the serum concentration of nineteen biomarkers from control, high risk, stage I-II, and stage III-IV woman. The sample size is indicated in each group column.

Biomarker	Normal risk N=65	High risk N=50	Stage I and II N=45	Stage III and IV N=39
VEGF	100[a]	95[a]	89[a]	97[a]
HGF	100	119	135	156
CA15-3	100	114	140	226
CEA	100	107	113	170
PSA	100	89	104	184
HER2	100	164	210	265
TNFα	100	83	79	71
E-selectin	100	96	105	110
MMP1	100	100	99	107
MMP2	100	99	101	106
MMP9	100	94	100	112
RANTES	100	96	96	108
sICAM1	100	56	94	111
IGF1	100	104	98	104
PDGF-AA	100	93	108	113
PDGF-BB	100	114	122	148
TGFα	100	66	75	54

[a] average percent with repect to normal risk average

Additionally, we have utilized this microarray ELISA to determine the presence or absence of these potential biomarkers in nipple aspirate fluid (NAF), a fluid that may be superior to serum for the detection of breast cancer. In pursuing this goal we first initiated a proteomic approach using 2-D column chromatography and mass spectrometry to identify proteins in NAF. Using this approach we were able to identify 63 NAF proteins, including at least 15 proteins that have been reported to be altered in serum or tumor tissue from women with breast cancer. This work was published in the journal *Breast Cancer Research and Treatment* and is included here as Appendix 2. We have done an initial experiment with the microarray ELISA and NAF. The results from this experiment are promising and we are pursuing them further.

Key Research Accomplishments

- Refinement of protein microarray resulting in a sensitive, quantitative, and reproducible assay.
- Demonstration of the utility of the microarray by comparing the concentration of serum HGF in woman with breast cancer and a healthy control group.
- Demonstated the ability to use the microarray for the simultaneous quantitation of multiple biomarkers.
- Standard curves for nineteen biomarkers generated.
- The simultaneous quantitation of nineteen different breast cancer biomarker levels from the serum of 100 normal and 100 breast cancer patients.
- Proteomic analysis of NAF identified 63 proteins.

Reportable Outcomes

Manuscripts

- Woodbury RL, Varnum SM, Zangar RC. Elevated HGF Levels in Sera from Breast Cancer Patients Detected Using a Protein Microarray ELISA. Journal of Proteome Research, 1, 233-237, 2002.
- Varnum SM, Covington, CC, Woodbury RL, Petritis, K, Kangas LJ, Abdullah, MS, Pounds JG, Smith RD, RC Zangar. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. Breast Cancer Res and Treatment 80: 87-97, 2003.
- Woodbury RL, Varnum SM, Zangar RC. Analysis of Multiple Breast Cancer Biomarkers in Human Serum Using a Protein Microarray ELISA. In preparation.

Book Chapters

- Varnum SM, RL Woodbury, and RC Zangar. 2003. "A Protein Microarray ELISA for Screening Biological Fluids." In Methods in Molecular Biology: Protein Arrays.

Presentations

- Woodbury RL, Varnum SM, Zangar RC. Elevated HGF Levels in Sera from Breast Cancer Patients Detected Using a Protein Microarray ELISA. Second Annual International Conference on Protein Microarrays, March 18-19, 2002.
- RC Zangar, R. Woodbury, S Varnum. "Development of a Protein Microarray ELISA for Measuring Cancer Biomarkers". Presented by Rick Zangar at Department of Defense's *Era of Hope* meeting, September 25-28, 2002, Orlando, Florida.
- Varnum S, Woodbury, R., R. Zangar. "A Protein Microarray ELISA for Measuring Cancer Biomarkers." Presented by Susan Varnum at PANWAT meeting, Richland, WA on September 19, 2002.
- Zangar RC, RL Woodbury, SM Varnum, CC Covington, and RD Smith. "Development of a Microarray ELISA for Characterizing Potential Markers of Breast Cancer in Nipple Aspirate Fluid." Presented by Richard C Zangar at Society of Toxicology, 2003 Annual Meeting, Salt Lake City, UT on March 10, 2003.
- Zangar RC, RL Woodbury, and SM Varnum. "Development of a user-friendly microarray ELISA for the analysis of potential protein markers of breast cancer." Presented by Rick Zangar at American Society of Biochemistry and Molelcular Biology, San Diego, CA on April 16, 2003.

- Varnum, S. "Proteomic Applications in Cancer Diagnostics and Viral Research". Invited speaker, University of Idaho on April 25, 2003.

Personnel

- Richard C. Zangar
- Susan M. Varnum
- Ronald L. Woodbury

Conclusions

We have developed a microarray ELISA capable of high-throughput analysis of potential breast cancer-specific biomarkers. The microarray ELISA was used to determine the serum concentration of these potential breast cancer-specific markers from 200 patients either with or without breast cancer. The serum concentrations for at least eleven of the potential markers did not alter significantly between the control groups and the patient groups diagnosed with breast cancer. However, the serum concentration for six biomarkers, hepatacyte growth factor, CEA, HER2, PSA, CA15-3 and TGF-α, were significantly altered between patients with breast cancer and control patients. This type of antibody microarray has great potential for the rapid determination of circulating markers associated with breast cancer. This basic technology is likely to lead to the development of more advanced chips with wide application in screening, diagnosis and prognosis of patients with breast cancer.

References

Appendices

Appendix 1 is an original copy of our journal article referenced in the text.
Appendix 2 is an original copy of our journal article referenced in the text.

research articles Journal of proteome research

Elevated HGF Levels in Sera from Breast Cancer Patients Detected Using a Protein Microarray ELISA

Ronald L. Woodbury, Susan M. Varnum, and Richard C. Zangar*

Battelle, Pacific Northwest National Laboratory, Molecular Biosciences Department, P7-56, P.O. Box 999, Richland, Washington 99352

Received January 30, 2002

We developed an ELISA in high-density microarray format to detect hepatocyte growth factor (HGF) in human serum. The microassay can detect HGF at sub-pg/mL concentrations in sample volumes of 100 μL or less. The microassay is also quantitative and was used to detect elevated HGF levels in sera from recurrent breast cancer patients. The microarray format provides the potential for high-throughput quantitation of multiple biomarkers in parallel, as demonstrated with a multiplex analysis of five biomarker proteins.

Keywords: microarray • ELISA • breast cancer • hepatocyte growth factor

Introduction

Enzyme-linked immunosorbent assay (ELISA)-based immunoassays have been the mainstay of the clinical laboratory for decades; however, problems arise when limited sample volume is available and high-throughput analysis of multiple markers is required. Protein microarrays potentially permit the simultaneous measurement of many proteins in a small sample volume and therefore provide an attractive alternative approach for the quantitative measurement of proteins in serum. To develop this potential, it is necessary that protein microarrays be both sensitive and quantitative and that they be available in a high-density format.

There have been several recent examples of the development and use of protein microarrays (reviewed in refs 1 and 2). Protein arrays have been used to screen the binding specificities of protein expression libraries[3] and for high-throughput screening of antibodies[4,5] and to examine protein–protein,[6–8] protein–DNA, and protein–RNA interactions.[9] Protein microarrays, in an ELISA-format, have also been developed for the measurement of proteins in clinical applications, for instance for the measurement of cytokines in conditioned media and serum,[10–12] prostate-selective antigen (PSA), PSA-ACT and IL-6 in serum,[13] and auto-antibodies in the sera of patients with autoimmune disease.[14]

Protein microarrays for the analysis of clinical samples need to be highly sensitive and quantitative. A variety of different surfaces have been used for making protein microarrays, including membranes, such as nitrocellulose and PVDF,[9,10,14] hydrogels,[15] glass,[6–8,16] and polystyrene.[17] In general, glass slides are the preferred surface for a microarray because of their ease of use, greater durability, optical properties, and the ability to use robotic spotters to generate high-density arrays. While a number of protein microarrays have been developed on glass slides, only a few have been developed for applications requiring high sensitivity. Sensitivities have ranged from 0.1 pg/mL to 1 ng/mL.[6,11,13,14,18] However, the most sensitive microarray developed (0.1 pg/mL), which utilizes the "rolling circle DNA amplification" technology,[18] requires extensive chemical labeling of the detection antibody and is not easily adaptable in other laboratories. Other sensitive assays require specialized equipment[11] or were developed for specific clinical applications such as the diagnosis of autoimmune disease and are not generally applicable.[14] As such, the development of a highly sensitive microarray ELISA that utilizes high-density spotting would advance this technology to a point where it is easily adaptable for high-throughput, quantitative analysis of proteins in clinical or research laboratory settings.

In this paper, we describe a microarray technology that is capable of the sensitive quantitation of hepatocyte growth factor (HGF), a protein recognized as a serum marker for a number of cancers, including breast cancer.[19] By coupling a microarray-ELISA format with the signal amplification of tyramide deposition, we obtain sub-pg/mL sensitivity. We demonstrate the utility of our microarray by comparing the concentration of serum HGF in women with breast cancer and a healthy control group and by showing that our results are comparable to those obtained with a commercial 96-well ELISA. This microarray is simple to prepare and highly sensitive and has the potential to be used to simultaneously analyze large numbers of serum proteins in a rapid and reproducible manner.

Experimental Section

Materials and Reagents. BS^3 and the protein biotinylation kit were from Pierce (Rockford, IL). HGF, HGF-specific, and vascular endothelial growth factor (VEGF)-specific antibodies, as well as the Quantikine ELISA kit for human HGF, were from R&D Systems (Minneapolis, MN). Other antibodies and purified marker proteins include the following: VEGF (Biodesign, Saco,

* To whom correspondence should be addressed. Phone: (509) 376-8596. Fax: (509) 376-6767. E-mail: richard.zangar@pnl.gov.

10.1021/pr025506q CCC: $22.00

Published on Web 04/18/2002

ME), CA 15-3 and anti-CA 15-3 antibodies (Fitzgerald, Concord, MA), soluble FAS ligand (Alexis Biochemicals, San Diego, CA), anti-FAS ligand antibodies (BD PharMingen, San Diego, CA), PSA and anti-PSA capture antibody (BiosPacific, Emeryville, CA), biotinylated anti-PSA antibody (Chromaprobe, Aptos, CA). The TSA Biotin System kit including blocking reagent, streptavidin–horseradish peroxidase (HRP) conjugate, biotinyltyramide, and reaction diluent was from Perkin-Elmer (Boston, MA). The Cy3–streptavidin conjugate was from Amersham Pharmacia (Piscataway, NJ). Sera from 10 breast cancer patients and 10 age-matched controls were obtained from the Breast Cancer Serum Biomarkers Resource, Lombardi Cancer Center (Washington, DC). Aminosilanated slides and all chemicals not listed above were obtained from Sigma (St. Louis, MO).

Microarray Preparation. A PixSys 5000 robot from Cartesian Technologies (Irvine, CA) equipped with ChipMaker2 quill pins from TeleChem (Sunnyvale, CA) was used to make the arrays. Aminosilanated slides were modified with 200 μL of a fresh 0.3 mg/mL solution of the homobifunctional cross-linker BS^3 in PBS (Dulbecco's phosphate buffered saline) for 5 min. The slides were rinsed briefly in 70% ethanol and dried under a stream of N_2 gas. An HGF-specific monoclonal "capture" antibody suspended to 1 mg/mL in PBS was printed on the slides. Also printed on each slide were an antibody that does not recognize HGF and a biotinylated protein. The antibody that does not recognize HGF served as a negative control. The biotinylated protein was a positive control for surface attachment and binding of the fluorescent probe (see below). The biotinylated protein also served as a reference when the array was imaged. These proteins were printed as arrays containing five spots of each reagent. Spots were printed either 0.5 or 1 mm apart and were approximately 1 nL in volume. The slides were incubated in a humid chamber for 1 h. Chamber humidity was maintained at 75% during all steps.

HGF Microassay. The arrays were circled with a hydrophobic pen to mark their location and to facilitate probing the array with small volumes. The pen makes a hydrophobic barrier on the surface of the slide, holding the sample in place over the array. During this step, the arrays were permitted to dry for 5–10 min. Each array was then blocked with 50 μL of TNB (100 mM Tris pH 7.5, 150 mM NaCl, 0.5% blocking reagent) for 1 h. The TNB was aspirated from the surface, and each array was incubated overnight with either 50 μL of an HGF standard in TNB or a serum sample diluted 4-fold in TNB (100 μL volumes were used in the high sensitivity experiment). The antigen solution was rinsed off in a gentle stream of water, and the slides were washed three times for 5 min in TNT (100 mM Tris pH 7.5, 150 mM NaCl, 0.05% Tween-20). Each array was then probed for 2 h with 50 μL of biotinylated detection antibody diluted in TNB. The biotinyl-anti-HGF antibody was diluted 1:1500 to 67 ng/mL for this step unless noted otherwise. Excess liquid was blotted from the slides, and the slides were washed three times for 5 min with TNT. The TSA–biotin system was then used to amplify the signal. Arrays were incubated for 1 h with 50 μL of streptavidin–HRP conjugate diluted 1:100 in TNB and washed as above. Each array was incubated for 10 min with 50 μL of biotinyltyramide diluted 1:100 in the supplied reaction diluent (or, alternatively, in 100 mM borate pH 8.5, 0.0009% H_2O_2), and the wash procedure was repeated. Each array was probed for 1 h in the dark with 50 μL of Cy3–streptavidin conjugate diluted to 1 μg/mL in TNB. Exposure to the light was avoided while the wash procedure was repeated, and the slides were rinsed twice in water and air-dried. A ScanArray 3000 from General Scanning (Billerica, MA) was used for fluorescence detection of the Cy3. Images thus captured in the ScanArray software were quantitated using ImaGene software (Biodiscovery). For comparison to our microarray ELISA, a commercial 96-well HGF ELISA was performed according to the manufacturer's instructions. Statistical comparison of the HGF levels in breast cancer patients and age-matched controls was undertaken using a **t** test and a probability value of <0.05 with SigmaStat 2.0 software.

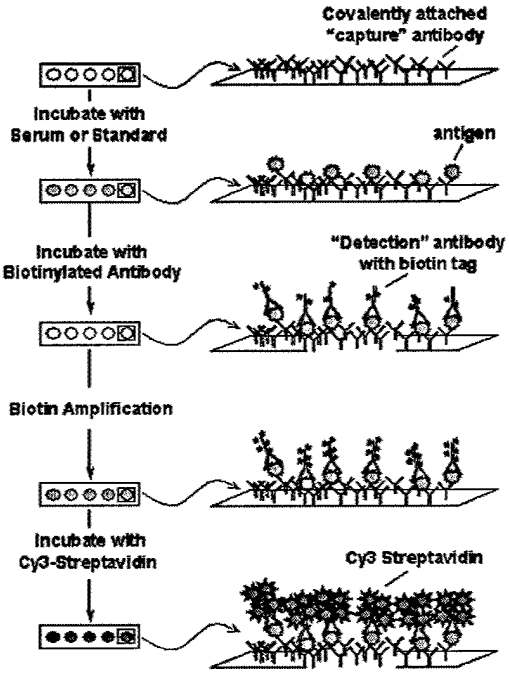


Figure 1. Schematic representation of the microarray "sandwich" ELISA used in this study.

Multiplex Experiment. This experiment was performed essentially as described for the HGF microassay. Capture antibodies for HGF, vascular endothelial growth factor, CA 15–3, FAS ligand, and PSA were spotted as solutions ranging from 0.25 to 1.0 mg/mL. Antigen concentrations were 200 pg/mL HGF, 300 pg/mL VEGF, 30 U/mL (approximately 60 ng/mL) CA 15-3, 200 pg/mL FAS ligand, and 20 pg/mL of PSA. Detection antibodies were used in concentrations ranging from 50 to 500 ng/mL. The CA 15-3 detection antibody was biotinylated using a kit and according to the manufacturer's (Pierce) instructions. All other detection antibodies were purchased as biotin conjugates. Two tyramide amplification steps were performed as described above. The first round of amplification was done after the arrays were exposed to detection antibodies for PSA and FAS ligand only. Subsequently, the arrays were exposed to the remaining detection antibodies and the amplification procedure repeated.

Results and Discussion

The sensitive detection of specific proteins is a major challenge in the development of protein microarrays designed to monitor levels of biomarkers that are often in low abundance. Since proteins cannot be amplified the way nucleic acids can, other methods of signal enhancement must be used if high levels of sensitivity are to be achieved. We have chosen to use an enzymatic signal enhancement method known as tyramide

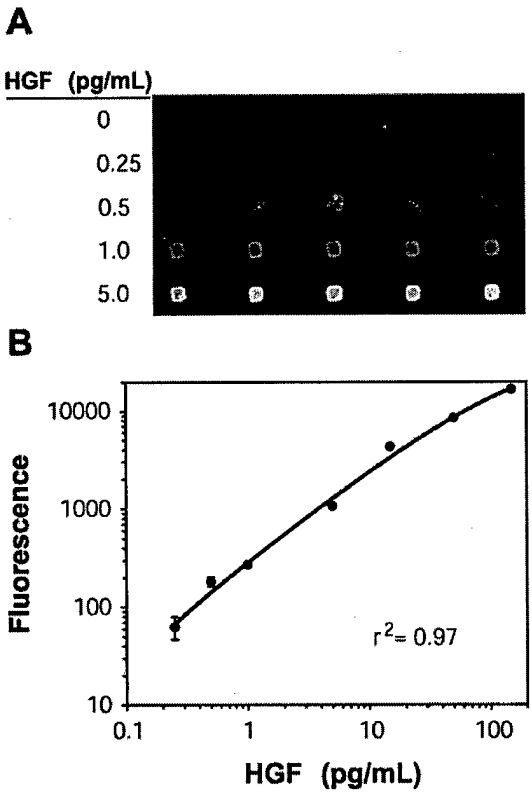


Figure 2. Microassay for HGF is capable of sub-pg/mL sensitivity. (A) The HGF concentration-dependent fluorescent response. Each row of five spots is from a separate array probed with the indicated HGF concentration. Images from separate microarrays are juxtaposed for comparison. (B) A standard curve for the HGF microarray values was calculated using a four-parameter logistic curve. Each data point was weighted by the inverse of the square of fluorescence intensity ($1/y^2$). Each data point represents the mean ± SE of five fluorescent spots for each HGF concentration.

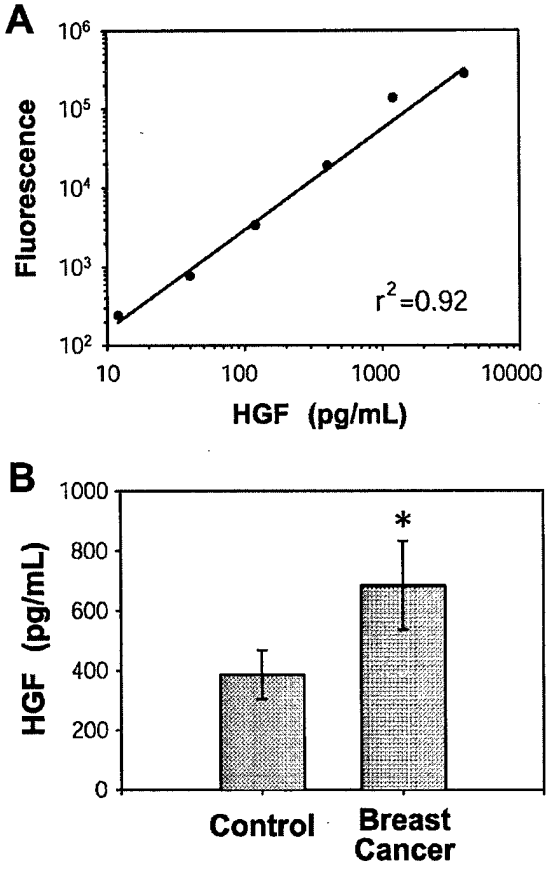


Figure 3. Detection of increased HGF levels in sera from breast cancer patients using the HGF microassay. (A) The HGF standard curve spans over 2.5 orders of magnitude. (B) HGF concentration (mean ± SE) in sera from breast cancer patients (**n** = 10) and normal controls (**n** = 10), as determined using the microassay. *Significantly different from the control group (**P** < 0.05).

signal amplification (TSA). This method has been used extensively in immunohistochemistry, a slide-based protein application, and has been found to provide exceptional sensitivity and low background. It has also been used in quantitative 96-well ELISA formats to detect specific proteins, such as HIV-1 p24 antigen and soluble interleukin 2 receptor, in complex body fluids.[20-22] Therefore, we tested tyramide signal amplification to see if it would be suitable for use with the microarray ELISA analysis.

A schematic diagram of the microarray ELISA approach used in this study is shown in Figure 1. Capture antibodies are covalently attached to a chemically reactive glass slide surface using spot sizes that are compatible with high-density microarrays. These spatially confined antibodies bind a specific antigen from a sample overlaying the array. A second, biotinylated antibody that recognizes the same antigen as the first antibody but at a different epitope is then used for detection. This "sandwich" approach favors specificity in analyte detection, since selective detection is provided sequentially by two separate antibodies. A streptavidin-HRP conjugate is then bound to the biotin moiety of the detection antibody, and catalyzes the TSA reaction. During this reaction the localized deposition of biotin takes place on the surface of all immediately available proteins. Thus the amount of covalently linked biotin in the immediate area is amplified. The biotin is then bound by a Cy3-streptavidin conjugate and the spot quantified using a fluorescence microarray reader. The amplification step does not decrease spot resolution as compared to spots of directly deposited proteins with fluorescent labels (data not shown).

We have successfully employed our microassay in the detection of HGF. By using a 1:200 dilution (0.5 μg/mL) of the detection antibody and 100 μL sample volumes, HGF can be detected down to 0.5 pg/mL (6 fM), equivalent to only 0.6 amol of HGF in the whole sample (Figure 2A). The quantitative range under these conditions approaches 3 orders of magnitude (Figure 2B). As we demonstrate below, we can manipulate the limits of the quantitative range by altering the concentration of the detection antibody. Antibodies that do not recognize HGF were printed as a negative control. The fluorescent intensity at the negative control spots in the presence of even the highest concentrations (1000 pg/mL) of HGF tested was comparable to the intensity of the spots containing anti-HGF capture antibody when incubated in solutions lacking HGF (data not shown). Since the same detection antibody was used in both cases, the low level of background fluorescence is not related to nonspecific binding of the detection antibody or HGF to the spots.

To measure HGF in clinical samples we sought to shift the quantitative range of the assay closer to the physiological range expected for HGF. By further diluting the detection antibody,

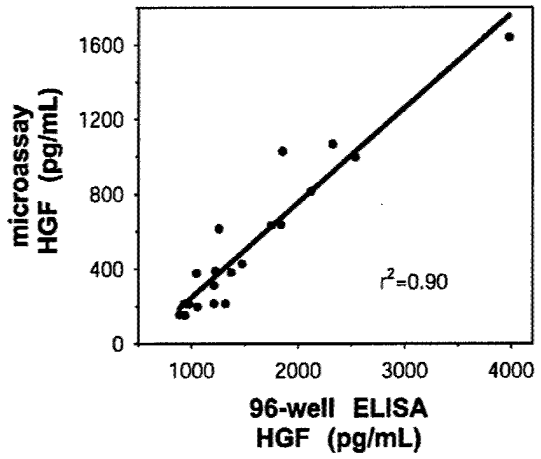


Figure 4. HGF values obtained with the microarray ELISA correlate well with a commercial 96-well ELISA. HGF concentration was measured by both methods in sera from 10 breast cancer patients and 10 age-matched controls.

we obtained a quantitative range from 12 to 4000 pg/mL in the serum (Figure 3A). Since serum samples were diluted 4-fold for this assay, each replicate of the microarray assay used only 12.5 μL of serum. HGF concentrations in clinical samples ranged from 0.15 to 1.64 ng/mL. Sera from 10 breast cancer patients with recurrent disease had a significantly elevated mean HGF concentration of 684 pg/mL (199−1640 pg/mL) compared to 386 pg/mL (153−998 pg/mL) in sera from 10 age-matched normal controls (Figure 3B). This result confirms previous work correlating recurrent breast cancer with higher levels of HGF in serum.[23]

To validate the results obtained with the microassay, we compared the data with that from a commercial 96-well ELISA kit. Data from the two methods showed a linear relationship with a correlation coefficient (r^2) of 0.90 (Figure 4), indicating that both methods produce similar results. Even so, the line describing the relationship between microassay and ELISA data does not have a slope of 1, meaning that the two assays give different absolute values for the HGF concentration in a given sample. It is common for assays based on immunochemical methods to vary in absolute quantitation, and improving their comparability is a recognized challenge.[24] Since the same set of standards but not the same antibodies were used in both assays, differences in results most likely reflect differences in antibody specificities, which may yield variable results due to steric interactions between the antibodies or differential recognition of the antigen due to post-translational modification or partial degradation of the antigen. This point is highlighted by a study in which six different ELISAs reported vastly different

Table 1. Interplate Reproducibility of the Multiplex Microarray ELISA

		signal		
antigen	pg/mL	mean	STD	% CV
HGF	200	2216	207	9.3
VEGF	300	37 831	4775	12.6
CA 15-3	60 000	15 450	1374	8.9
FAS ligand	200	33 092	2591	7.8
PSA	20	23 515	947	4.0

concentrations of tumor necrosis factor in the majority of individual samples.[25] Despite differences between the microassay and the 96-well ELISA, the range of HGF concentrations we found in the sera of breast cancer patients using our microassay (0.199−1.64 ng/mL) is nearly identical to the range found by Maemuro and co-workers (0.15 to 1.43 ng/mL) using a different ELISA kit.[19]

To determine if this technology could be used to simultaneously detect multiple biomarkers, we analyzed five different proteins in a single microarray. The capabilities of the microarray were further tested by analyzing proteins over a wide range of concentrations. The proteins were HGF, VEGF, CA 15-3, FAS ligand, and PSA and were assayed at biologically relevant concentrations[26−32] that ranged from 20 to 60 000 pg/mL (Table 1). Furthermore, we only tested a single antibody pair for each protein. The goal here was to see if it was possible to modify assay conditions to accommodate varying antibody affinities and antigen levels. This approach is more efficient than testing different antibody combinations and may be essential when antibody availability is limited. Initial studies indicated that we could readily detect VEGF and CA 15−3 using incubation and detection conditions similar to those used for HGF, but that FAS ligand and PSA signals were very weak (data not shown). In an effort to increase signal strength, we tried using two tyramide amplification steps for these latter two antigens. In this procedure, the microarray was first incubated with the antigen mixture followed by incubation with biotinylated antibodies to FAS ligand and PSA. The biotin was then amplified using the tyramide deposition procedure. Then the microarray was incubated with a mixture of biotinylated antibodies against the remaining 3 antigens (i.e., HGF, VEGF, and CA 15-3). The subsequent tyramide amplification step would therefore be a second amplification for FAS ligand and PSA but would be the first amplification step for HGF, VEGF, and CA 15-3. Replicate microarray assays were then undertaken using this procedure such that three microarrays were exposed to all five antigens, while individual microarrays were prepared where individual antigens were omitted (Figure 5). Using this approach, we were clearly able to obtain usable signal/background levels for each biomarker (Figure 5). Analysis of

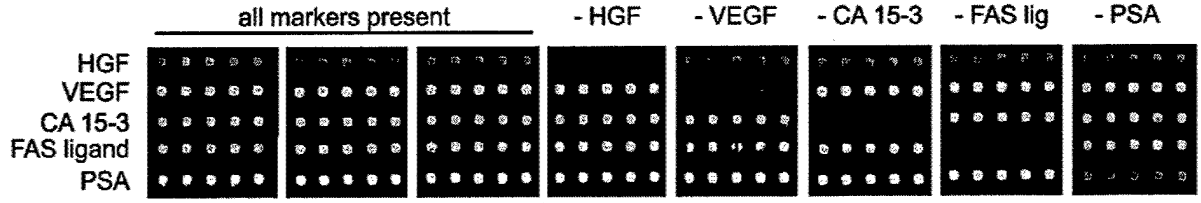


Figure 5. Multiple biomarkers can be simultaneously measured on a single microarray. Eight identical slides were printed with capture antibodies to five different protein markers. Three of these slides were incubated with a mixture of all five antigens (see Table 1), while the other five slides were each incubated with the same mixture of proteins minus a single antigen. Antigens were then detected as described in the text.

the coefficient of variation (CV; the standard deviation divided by the mean) between slides in the multiplex study indicated that the CV values varied from 4 to 12.6% (Table 1). Therefore, these data demonstrate that the microarray ELISA can be easily adapted for the reproducible analysis of multiple antigens, even when the concentrations of the different antigens vary 3000-fold, and there are apparent variations in the quality of the antibodies.

The microassay we describe has many advantageous features, including its small size, sensitivity, and the commercial availability of all reagents and detection equipment. The small size will allow for multiple biomarkers to be analyzed in parallel. That is, with the spot size of ~150 μm that we used, it is possible to make high-density microarrays with 5000−10000 spots per slide. Small size also translates into more efficient use of reagents and precious biological samples such as biopsies or nipple aspirates. Exceptional sensitivity and flexible quantitative range increases the pool of biomarkers that can potentially be assayed, both individually and simultaneously. As such, the ability to vary the quantitative range of individual biomarkers simply by varying the concentration of their respective detection antibodies should prove particularly useful for assaying multiple protein markers on a single microarray. The assay can be prepared and run without the need for customized detection equipment or in-house protein modification, which will facilitate the rapid development and use of similar microarrays in other laboratories.

Conclusions

We developed a protein microarray ELISA suitable for analysis of HGF levels in serum samples. This assay demonstrated exceptional sensitivity and quantitative characteristics comparable with a 96-well ELISA. This technology is readily adaptable for high-throughput, high-density analysis of proteins in clinical and research laboratories.

Acknowledgment. This project was sponsored by the Department of the Army's Breast Cancer Research Program, award no. DAMD17-00-1-0132. The content of this work does not necessarily reflect the position or the policy of the U.S. Government, and no official endorsement should be inferred.

References

(1) Cahill, D. J. **J. Immunol. Methods 2001**, **250**, 81−91.
(2) Kodadek, T. **Chem. Biol .2001**, **8**, 105−115.
(3) Lueking, A.; Horn, M.; Eickhoff, H.; Bussow, K.; Lehrach, H.; Walter, G. **Anal. Biochem. 1999**, **270**, 103−111.
(4) Holt, L. J.; Bussow, K.; Walter, G.; Tomlinson, I. M. **Nucl. Acids Res. 2000**, **28**, E72.
(5) de Wildt, R. M.; Mundy, C. R.; Gorick, B. D.; Tomlinson, I. M. **Nat. Biotechnol. 2000**, **18**, 989−994.
(6) Haab, B. B.; Dunham, M. J.; Brown, P. O. **Genome Biol. 2001**, **2**.
(7) MacBeath, G.; Schreiber, S. L. **Science 2000**, **289**, 1760−1763.
(8) Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M. **Science 2001**, **293**, 2101−2105.
(9) Ge, H. **Nucleic Acids Res. 2000**, **28**, e3.
(10) Huang, R. P.; Huang, R.; Fan, Y.; Lin, Y. **Anal. Biochem. 2001**, **294**, 55−62.
(11) Huang, J. X.; Mehrens, D.; Wiese, R.; Lee, S.; Tam, S. W.; Daniel, S.; Gilmore, J.; Shi, M.; Lashkari, D. **Clin. Chem. 2001**, **47**, 1912−1916.
(12) Huang, R. P. **Clin. Chem. Lab. Med. 2001**, **39**, 209−214.
(13) Wiese, R.; Belosludtsev, Y.; Powdrill, T.; Thompson, P.; Hogan, M. **Clin. Chem. 2001**, **47**, 1451−1457.
(14) Joos, T. O.; Schrenk, M.; Hopfl, P.; Kroger, K.; Chowdhury, U.; Stoll, D.; Schorner, D.; Durr, M.; Herick, K.; Rupp, S.; Sohn, K.; Hammerle, H. **Electrophoresis 2000**, **21**, 2641−2650.
(15) Arenkov, P.; Kukhtin, A.; Gemmell, A.; Voloshchuk, S.; Chupeeva, V.; Mirzabekov, A. **Anal. Biochem. 2000**, **278**, 123−131.
(16) Mendoza, L. G.; McQuary, P.; Mongan, A.; Gangadharan, R.; Brignac, S.; Eggers, M. **Biotechniques 1999**, **27**, 778−780, 782−786, 788.
(17) Ekins, R. P. **Clin. Chem. 1998**, **44**, 2015−2030.
(18) Schweitzer, B.; Wiltshire, S.; Lambert, J.; O'Malley, S.; Kukanskis, K.; Zhu, Z.; Kingsmore, S. F.; Lizardi, P. M.; Ward, D. C. **Proc. Natl. Acad. Sci. U.S.A. 2000**, **97**, 10113−10119.
(19) Maemura, M.; Iino, Y.; Yokoe, T.; Horiguchi, J.; Takei, H.; Koibuchi, Y.; Horii, Y.; Takeyoshi, I.; Ohwada, S.; Morishita, Y. **Cancer Lett 1998**, **126**, 215−220.
(20) Gotoh, Y.; Okamoto, Y.; Uemura, O.; Mori, N.; Tanaka, S.; Ando, T.; Nishida, M. **Clin. Chim. Acta 1999**, **289**, 89−97.
(21) Boni, J.; Opravil, M.; Tomasik, Z.; Rothen, M.; Bisset, L.; Grob, P. J.; Luthy, R.; Schupbach, J. **AIDS 1997**, **11**, F47−52.
(22) Schupbach, J.; Boni, J.; Flepp, M.; Tomasik, Z.; Joller, H.; Opravil, M. **J. Med. Virol. 2001**, **65**, 225−232.
(23) Taniguchi, T.; Toi, M.; Inada, K.; Imazawa, T.; Yamamoto, Y.; Tominaga, T. **Clin. Cancer Res. 1995**, **1**, 1031−1034.
(24) Bidart, J. M.; Thuillier, F.; Augereau, C.; Chalas, J.; Daver, A.; Jacob, N.; Labrousse, F.; Voitot, H. **Clin. Chem. 1999**, **45**, 1695−1707.
(25) de Kossodo, S.; Houba, V.; Grau, G. E. **J. Immunol. Methods 1995**, **182**, 107−114.
(26) Black, M. H.; Giai, M.; Ponzone, R.; Sismondi, P.; Yu, H.; Diamandis, E. P. **Clin. Cancer Res. 2000**, **6**, 467−473.
(27) Borchert, G. H.; Melegos, D. N.; Tomlinson, G.; Giai, M.; Roagna, R.; Ponzone, R.; Sgro, L.; Diamandis, E. P. **Br. J. Cancer 1997**, **76**, 1087−1094.
(28) Canizares, F.; Sola, J.; Perez, M.; Tovar, I.; De Las Heras, M.; Salinas, J.; Penafiel, R.; Martinez, P. **Tumour Biol. 2001**, **22**, 273−281.
(29) Heer, K.; Kumar, H.; Read, J. R.; Fox, J. N.; Monson, J. R.; Kerin, M. J. **Clin. Cancer Res. 2001**, **7**, 3491−3494.
(30) Heinze, T.; Schurenkamper, P.; Minguillon, C.; Lichtenegger, W. **Anticancer Res. 1997**, **17**, 2953−2954.
(31) Salven, P.; Perhoniemi, V.; Tykka, H.; Maenpaa, H.; Joensuu, H. **Breast Cancer Res. Treat. 1999**, **53**, 161−166.
(32) Tomokuni, A.; Otsuki, T.; Isozaki, Y.; Kita, S.; Ueki, H.; Kusaka, M.; Kishimoto, T.; Ueki, A. **Clin. Exp. Immunol. 1999**, **118**, 441−444.

PR025506Q

Breast Cancer Research and Treatment **80:** 87–97, 2003.

Report

Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer

Susan M. Varnum[1], Chandice C. Covington[2], Ronald L. Woodbury[1], Konstantinos Petritis[1], Lars J. Kangas[1], Mohamed S. Abdullah[3], Joel G. Pounds[1], Richard D. Smith[1], and Richard C. Zangar[1]

[1]*Pacific Northwest National Laboratory, Richland, WA;* [2]*University of California School of Nursing, Los Angeles, CA, USA;* [3]*National Health Research & Development Center, Nairobi, Kenya*

Key words: biomarkers, breast cancer, nipple aspirate fluid, proteomics

Summary

Mammary ductal cells are the origin for 70–80% of breast cancers. Nipple aspirate fluid (NAF) contains proteins directly secreted by the ductal and lobular epithelium in non-lactating women. Proteomic approaches offer a largely unbiased way to evaluate NAF as a source of biomarkers and are sufficiently sensitive for analysis of small NAF volumes (10–50 μl). In this study, we initially evaluated a new process for obtaining NAF and discovered that this process resulted in a volume of NAF that was suitable for analysis in ~90% of subjects. Proteomic characterization of NAF identified 64 proteins. Although this list primarily includes abundant and moderately abundant NAF proteins, very few of these proteins have previously been reported in NAF. At least 15 of the NAF proteins identified have previously been reported to be altered in serum or tumor tissue from women with breast cancer, including cathepsin D and osteopontin. In summary, this study provides the first characterization of the NAF proteome and identifies several candidate proteins for future studies on breast cancer markers in NAF.

Introduction

Breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer deaths in women in the United States [1]. In contrast to most cancers, the incidence of breast cancer has been increasing in recent years [1]. Most breast cancer deaths are caused by metastatic disease, highlighting the importance of early detection and screening. However, existing detection methodologies have major shortcomings [2, 3]. None of the available screening technologies can distinguish breast cancer from benign breast disease and sometimes even normal breast tissue, resulting in a high rate of false-positive and false-negative reports [4]. For instance, mammography only detects cancer in 70–90% of individuals with the disease, while the rate of false positives is from 5 to 17% [5, 6]. Additionally, current prognostic procedures are poor at detecting micrometastases or early recurrent disease. Therefore, it is clear that new, non-invasive methods are needed to complement current methodologies for the detection and prognosis of precancerous and cancerous breast lesions when they are small and more easily treated.

Within the ductal and lobular system of the breast is a fluid that is continuously secreted and reabsorbed in non-pregnant/non-lactating women [7]. With the assistance of a gentle aspiration device, this breast fluid can be extracted through the nipple and is referred to as nipple aspirate fluid (NAF) [8]. NAF potentially offers a superior fluid for detection of breast cancer than serum since the proteins present are specifically from breast tissue. This fluid collects from the epithelial cells lining the ductal system of the breast, the same cells that are the source of 70–80% of breast cancers. Therefore, it is not surprising that NAF has been found to be a rich source of breast cancer biomarkers [9].

We recently developed a very sensitive, high-density microarray ELISA that is suitable for high-throughput, quantitative analysis of hundreds of

proteins in small-volume samples such as NAF [10]. As such, it is now possible to rapidly evaluate a large number of NAF proteins for their utility as markers of breast cancer. Unfortunately, only limited studies have been undertaken characterizing the protein content of NAF. For example, a recent biochemical characterization of NAF only identified 10 proteins or associated enzymatic activities [11]. Another study identified ~50–100 spots by two-dimensional gel electrophoresis, but the majority of the spots appeared to represent multiple glycosylation states of five or six highly abundant, unidentified proteins [12]. Therefore, a better characterization of the proteins present in NAF is needed to identify candidate proteins for examination as cancer markers.

Improvements in nipple aspirators have yielded devices that generally provide greater NAF volume. Similarly, advances in 'gel-free' proteomic technologies based upon mass spectrometry now allow for the identification of proteins in very small samples. In this study, we utilize a specialized, non-invasive, well-tolerated aspirator and improved aspiration process [13, 14] to obtain NAF samples of high quality. We demonstrate that NAF is a highly concentrated source of protein that is well suited for analysis of cancer markers. We also identify 15 proteins that have been reported to be potential markers for breast cancer in serum or tumor tissue but have not previously been identified in NAF.

Materials and methods

NAF collection and processing. NAF samples were collected from women in the Midwest United States and from a rural region in Kenya. NAF samples were collected and analyzed with approval by the Wayne State University and the Pacific Northwest National Laboratory Institutional Review Boards. Two separate sources for the NAF samples were due to availability from other studies rather than for scientific purposes associated with this study. Therefore, no comparisons were performed between the two sources. Donors ($N = 121$) were apparently healthy women, who were not currently pregnant (as determined by a pregnancy test in women with current menses), taking exogenous hormones, or lactating, and were at least 35 years of age or older. The Kenyan women experienced youthful multiparity and extensive lactation histories. The American women were selected from an existing volunteer pool of urban-residing women previously identified by the Community Outreach Core through Karmanos Cancer Institute in Detroit, Michigan and through flyers and radio announcements. Of this Midwest group ($N = 75$), 62 were Caucasian, 12 were African American, and 1 was Asian. The women ranged in age from 35 to 70, with a mean age of 47. Overall, from both donor groups, 57 women were peri- ($N = 29$) or post-menopausal ($N = 28$). Additionally, 17 of the women in the Detroit group and none in the Kenyan group reported benign breast lumps that had been diagnosed by needle biopsy. These women were not excluded from the study. At the beginning of the research clinic visit, the Morrow and American Cancer Association clinical breast examination method was performed to detect the presence of 'lumps', fibrocystic changes, and other breast conditions [15]. If no suspicious breast findings were detected, then following venipuncture, the women began the procedure for collecting NAF, which is a clinical intervention process, starting with an initial attempt for the women to self-aspirate nipple fluid. This process started with the woman using glycerin to conduct a 5 min breast massage. Small body heating pads were placed on the breast and held in place with a sports-type bra for 15 min. Nipple fluids were then aspirated bilaterally using a patented NAF collection system developed by Dr Covington (NeoMatrix, Irvine, CA). The NAF was collected in a capillary tube, transferred to a microtube, wrapped in foil, and stored at −70°C until analysis. NAF protein concentration was determined as described previously [16]. Since NAF samples typically are very viscous, the samples were first diluted in 50–400 µl of phosphate-buffered saline (PBS), pH 7.4, depending on the initial sample volume, and vortexed to improve handling. Soluble NAF proteins were isolated from particulate material and a buoyant lipid layer by centrifugation at 14,000 × *g* for 1 min.

Overview of the proteomic analyses of NAF. NAF was analyzed using three methodologies. In the first analysis, major proteins in NAF were analyzed by in-gel digestion. Once identified, these abundant proteins were removed from the NAF sample by affinity chromatography. Since the total peptide mass that we loaded onto the capillary liquid chromatography (cLC) mass spectrometer was limited to ~5–10 µg, removal of abundant peptides effectively increased the mass of lower-abundance peptides that could be analyzed in each MS run in the subsequent two analyses. In the second analysis, NAF peptides were analyzed by cLC-tandem MS without prior fractionation of the

peptides. In order to further increase the number of low-abundance peptides selected for analysis in each run, multiple cLC-tandem MS runs were performed using the same sample but the peptides selected for tandem MS during each run were restricted to a 200 *m/z* range. In the third analysis, peptides were fractionated by ion-exchange chromatography prior to analysis of individual fractions by cLC-tandem MS. This fractionation step served to concentrate individual peptides and simplify the peptide mixture in each MS run, thereby generally improving the quality of the tandem MS data obtained for peptides from low-abundance proteins. Each of these procedures is described in more detail below.

Identification of abundant proteins in NAF. NAF was fractionated by electrophoresis on a denaturing polyacrylamide gel containing lauryl sulfate, as described [17]. The four major protein bands present in Coomassie (GelCode Blue, Pierce Chemical Co., Rockford, IL) stained gels were excised and analyzed by in-gel trypsin digestion and tandem MS, as described [17]. Additionally, the four major protein bands were quantitated by densitometry individually as a percent of the total protein present in the gel lane using an image captured with overhead lighting on a Lumi-Imager (Boehringer Mannheim, Germany) and with LumiAnalyst Imaging software.

Removal of abundant proteins in NAF by affinity chromatography. Immunoglobulins were first removed from 5 mg of NAF protein by mixing with protein A/G and then protein L affinity columns. Specifically, the protein A/G beads (Pierce Chemical Co., Rockford, IL) were first equilibrated with 20 mM sodium phosphate (pH 8.0), and then 5 mg of NAF protein was incubated with the beads for 2 h at room temperature. The protein A/G treated fraction was collected by loading the beads into a column and washing with equilibration buffer. This fraction was further purified by applying it to a Protein L column (Pierce) diluted 1:1 in PBS. The column was washed with PBS and the immunoglobulin-depleted fraction was collected.

To prepare the albumin and lactoferrin affinity chromatography column, the relevant antibodies, either anti-albumin (Capricorn Products, Inc., Scarborough, ME) or anti-lactoferrin (Accurate Chemical Co., Westbury, NY) were covalently linked to UltraLink Biosupport Medium (Pierce Chemical Co., Rockford, IL) following the manufacturer's directions. The anti-human serum albumin (HSA) and the anti-lactoferrin affinity beads were combined to make a bed volume of 800 µl. The immunoglobulin-depleted NAF was combined with the anti-albumin and anti-lactoferrin beads in PBS and incubated overnight at 4°C with gentle mixing. The HSA and lactoferrin-depleted fraction was collected by generating a column with the beads and washing with equilibration buffer. This was followed by elution of the HSA and lactoferrin fraction with 0.2 M glycine (pH 2.5).

In-solution tryptic digestion of NAF. NAF that was depleted of immunoglobulins, HSA, and lactoferrin was dialyzed against 100 mM ammonium bicarbonate. Proteins were denatured by addition of urea to 8 M and heating to 37°C for 30 min. The sample was then diluted 4-fold with 100 mM ammonium bicarbonate and $CaCl_2$ was added to 5 mM. Methylated, sequencing-grade trypsin (Promega, Madison, WI) was added at a substrate-to-enzyme ratio of 20:1 (mass:mass) and incubated at 37°C for 15 h.

Strong cation exchange separation of NAF peptides. Two hundred micrograms of NAF that was depleted of abundant proteins was dialyzed against 100 mM AB, lyophilized to dryness and trypsin digested as described above. Strong cation exchange chromatography was performed on the peptide sample utilizing a Synchropak S 300, 100 × 2 mm chromatographic column (Thermo Hypersil-Keystone, Bellefonte, PA, USA). A 1 h gradient was utilized at a flow rate of 200 µl/min with fractions collected every 2 min. The beginning solvent system was 25% acetonitrile, 75% water containing 10 mM $HCOONH_4$, pH 3.0, adjusted with formic acid, and the ending solvent system was 25% acetonitrile, 75% water containing 200 mM $HCOONH_4$, pH 8.0. The peptide mixture was resuspended in 25% acetonitrile, 75% water containing 10 mM $HCOONH_4$, pH 3.0 with formic acid prior to injection. Fractions were lyophilized and stored at −20°C until MS analysis.

Tandem mass spectrometric analysis of peptides. Peptide samples were analyzed by reversed phase cLC coupled directly with electrospray tandem mass spectrometers (Thermo Finnigan, models LCQ Duo and DecaXP). Chromatography was performed on a 60 cm, 150 µm i.d. × 360 µm o.d capillary column (Polymicro Technologies, Phoenix, AZ) packed with Jupiter C_{18} 5 µm-diameter particles (Phenomenex, Torrence, CA). A solvent gradient was used to elute the peptides

using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The gradient was linear from 0 to 5% solvent B in 20 min, followed by 5–70% solvent B in 80 min, and then 70–85% solvent B in 45 min. Solvent flow rate was 1.8 μl/min.

The capillary LC system was coupled to a LCQ ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) using an in-house manufactured ESI interface in which no sheath gas or makeup liquid was used. The temperature of heated capillary and electrospray voltage was 200°C and 3.0 kV, respectively. Samples were analyzed using the data-dependent MS/MS mode over the *m/z* range of 300–2000, 500–700, 675–875, 850–1050, 1025–1225, 1200–1400, 1375–1575, 1550–1750, and 1725–2000. The three most abundant ions detected in each MS scan were selected for collision-induced dissociation.

Sequest analysis. The SEQUEST algorithm [18] was run on each of the data sets against a modified version of the human.fasta from the National Center for Biotechnology Information. Modifications to the database included the removal of viral proteins and redundant protein entries. A peptide was considered to be a match by using a conservative criteria set developed by Yates and coworkers [19, 20]. Briefly, all accepted SEQUEST results had a delta Cn of 0.1 or greater. Peptides with a +1 charge state were accepted if they were fully tryptic and had a cross-correlation (X_{corr}) of at least 1.9. Peptides with a +2 charge state were accepted if they were fully tryptic or partially tryptic and had an X_{corr} of at least 2.2. Peptides with +2 or +3 charge states with an X_{corr} of at least 3.0 or 3.75, respectively, were accepted regardless of their tryptic state. When a protein was identified by two or fewer unique peptides that met the SEQUEST criteria above, the SEQUEST spectra alignment was manually validated using criteria described [20].

Prediction of peptide elution times using an artificial neural network. As an additional criteria for evaluating the quality of tandem MS data, an artificial neural network has been developed for predicting the elution time of peptides separated by reverse phase HPLC prior to on-line identification by MS [21]. In order to account for day-to-day and column-to-column variation, peptide elution times were normalized to a scale of 0–1 by using a genetic algorithm. Development of the neural network model was based on the amino acid composition of the peptides, using a dataset of ~7000 peptides that were identified with a high level of confidence. Application of this model to 5200 different peptides (also identified with a high level of confidence) produced a mean accuracy of ~3% and was able to distinguish a subset of peptides that were previously misidentified. As such, peptides in this study that were accepted based on the criteria described above (i.e., SEQUEST parameters and manual spectra examination) were further evaluated based on predicted and measured normalized elution times.

Results

The nipple aspiration system used here combined with the gentle massage and warming protocol prior to sample collection resulted in a high aspiration rate for NAF collection. Within the Detroit donors, 85% were able to aspirate NAF. All but one of the Kenyan woman could aspirate NAF (98%), resulting in an overall success rate of about 91%.

Our initial concern with NAF was that the small sample volumes (typically 10–50 μl) would make proteomic analysis impractical. Therefore, we undertook a preliminary characterization of two NAF samples obtained from women in Kenya and two from women in the United States. The protein concentrations in these four samples were exceptionally high, ranging from 45 to 200 mg/ml. Further analysis of the NAF samples on a denaturing SDS-PAGE gel indicated that although the samples varied in protein concentration, four major bands were observed in all samples in approximately equal proportions (data not shown). Combined, these results suggest that normalization of NAF samples against total protein content would be a reasonable approach for future studies designed to provide a quantitative analysis of biomarkers levels. This conclusion is consistent with a study showing marked changes in PSA levels in NAF samples from breast cancer patients when levels of PSA were normalized against total protein concentrations [22].

In order to further determine if there is sufficient protein in the NAF samples for proteomic analysis, we pooled 10 samples obtained from Kenyan donors. It should be noted that these were numerically the first 10 samples from a single study and were not preferentially selected because of large volume or any other sample characteristics. As such, they are likely to be representative of NAF samples in general. Analysis of the protein content indicated that the pooled samples contained 30 mg of protein. Therefore, the average NAF sample from Kenya contained 3 mg of

total protein. We then analyzed 36 individual samples obtained from women in the United States. The total protein content of these samples was 2.2 ± 0.5 mg (mean $\pm$ SE). The median protein content was 1.3 mg with values ranging from 0.22 to 14 mg. Since a typical LCQ tandem MS analysis only requires 5–10 μg of protein, these data indicate that there is more than sufficient protein content in the NAF samples for proteomic analysis even with the four most abundant proteins removed.

We used in-gel trypsin digestion and cLC-tandem MS to identify proteins in four dark-staining bands observed in NAF samples (Figure 1). We identified the most abundant proteins in the NAF samples as immunoglobulins, poly-immunoglobulin receptor, albumin and lactoferrin. Rough estimates of the abundance of the protein or group of proteins, estimated as a percent of the total protein mass in a pooled NAF sample, are 10% for poly-immunoglobulin receptor, 10% for lactoferrin, 15% for albumin and 30% for immunoglobulins. These values were derived based on the staining intensity of the bands on a stained SDS gel both before and after removal of a specific protein by affinity chromatography.

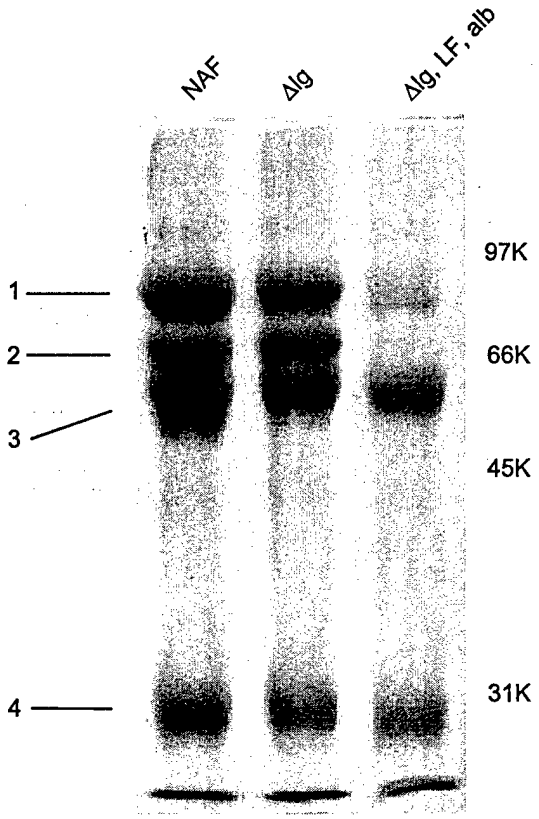


Figure 1. Removal of abundant proteins from NAF. A pooled NAF sample was subjected to affinity chromatography to remove abundant proteins. The original sample and the samples formed by depletion of only the immunoglobulins (ΔIg) and then also lactoferrin and albumin (ΔIg, LF, alb) were separated by SDS-PAGE and stained. Band 1 corresponds to polyimmunoglobulin receptor and lactoferrin, band 2 corresponds to albumin and bands 3 and 4 correspond to the immunoglobulin heavy and light chains, respectively.

A more detailed proteomic characterization of NAF was undertaken. Affinity chromatography was used to remove the abundant proteins prior to MS analysis. It is possible that the affinity chromatography steps could also deplete low abundance proteins that remain bound to the targeted abundant proteins even after extensive washing. However, since the total mass of peptides that could be analyzed in a single MS run was limited, the removal of the most abundant peptides effectively increased the mass of lower abundance proteins that could be analyzed in each run. Therefore, the affinity chromatography steps should increase sensitivity for most low-abundance proteins. A pooled NAF sample from Kenya was used as the starting material for this study. We have repeated these studies using a pooled NAF sample from the US and obtained similar results to the Kenya samples (results not shown). Using the pooled NAF sample from Kenya, immunoglobins were first removed from 5 mg of protein using protein A/G beads. This step was only partially effective, in that only ~30% of the immunoglobulin fraction was removed as determined by densitometry (Figure 1, compare lanes 1 and 2). This result is in contrast to serum, where the same procedure removed essentially all of the immunoglobulins [23]. The immunoglobulins remaining in the NAF sample were identified by in-gel digestion and LCQ tandem MS as IgM and IgA, which are not efficiently bound by protein A/G. Protein L beads, which bind IgM and IgA, were used and found to reduce the immunoglobulin fraction 20% as determined by densitometry (data not shown). Passage of the immunoglobin-depleted sample over beads containing antibodies specific for albumin and lactoferrin essentially removed all of both proteins from the NAF sample, as indicated in the stained gel (Figure 1, compare lanes 1 and 3) and subsequent MS analysis of the affinity-purified NAF eluant. Following depletion of abundant proteins, the NAF sample was denatured in urea and trypsin digested. Peptides were analyzed either without further processing (except buffer

Table 1. Summary of proteins identified in NAF

Protein name	NCBI accession no.	Number of unique peptides[a]	X_{corr} (Max)[b]	Potential biomarker[c]	Milk or immune protein	Glycosylated[d]	Membrane[e]
14-3-3 protein β/α	4507949	1	3.3				
14-3-3 protein zeta	4507953	1	2.2				
α_2-Actin	1070613	3	3.1				
β-Actin	4501885	4	3.2				
Albumin[f]	4502027	6	3.4		+	+	
α_1-Antichymotrypsin	2144574	15	4.5	+	+	+	
α_1-Antitrypsin[f]	1703025	13	4.5	+	+	+	
Aminopeptidase N	113743	4	5.3			+	T2
Antigen p97-melanotransferrin	5174559	1	2.9			+	T1
Apolipoprotein D	1246096	10	4.9	+		+	
Apolipoprotein E	4557325	7	3.5			+	
Butyrophilin	4502475	1	2.3			+	T1
Cancer-associated serine protease-protecting peptide (CRISPP peptide)	262839	1	3.3				
α-Casein	4503085	4	5.1		+		
β-Casein	256274	26	5.8		+	+	
κ-Casein	1705606	4	4.3		+	+	
Cathepsin D	4503143	3	5.1	+	+	+	
Ceruloplasmin (ferroxidase)	1070458	3	4.3	+	+	+	
CD14	115956	3	3.7		+	+	GPI
CD34 (prominin)	5174387	4	3.9			+	5TM
Clusterin	4502905	13	4.9	+	+	+	
Coagulation factor II	4503635	1	2.4		+		
Collagen α1	1360676	1	2.8			+	
Complement C3	4557385	8	5.6		+	+	
Complement C4	116602	19	4.9		+	+	
Complement C7	4557387	1	2.7		+	+	T2
Complement factor B	584908	3	3.3		+	+	
Complement factor D	1827805	1	3.1		+	+	
Enhancer protein	2135068	1	2.4				
Ephrin	4826708	1	2.3			+	
Fibrinogen gamma-B chain	120146	1	4.2	+	+	+	
α-L-fucosidase	2118051	1	2.2		+		
Gelsolin	4504165	4	3.3	+	+		

α_{1B}-Glycoprotein	112892	1	3.6			+	
α_2-Glycoprotein (Zn)	105274	17	5.9	+		+	
α_2-HS-glycoprotein	4502005	1	2.7			+	
Haptoglobin	123507	1	2.4		+	+	
Ig alpha chain (heavy)[f]	113584	19	5.1		+	+	
Ig gamma chain (heavy)	260394	2	2.2		+	+	
Ig mu chain (heavy)	106378	11	4.3		+	+	
Ig J chain	400044	5	3.7		+	+	
Ig kappa chain (light)[f]	106528	40	5.7		+	+	
Ig lambda chain (light)[f]	106631	17	4.2		+	+	
Iroquois-class homeodomain protein	5031807	1	3.3				
Keratin 1	5174491	2	3.7				
Keratin 2a	4557703	1	2.3				
α-Lactalbumin	4504947	2	3.2	+	+		
Lactoferrin[f]	6175096	24	5.0		+	+	
Leucine-rich α-2-glycoprotein	112908	1	2.1			+	
β-2-Microglobulin, *pI* 5.3	1195503	2	2.5		+		
Osteopontin	129260	7	4.3	+	+	+	
Polymeric-immunoglobulin receptor[f]	1730570	25	5.9		+	+	T1
Prolactin-induced protein (GCDFP-15)	4505821	4	3.8	+	+	+	
Prostasin (serine protease 8)	4506153	1	2.6			+	T1
Pyruvate kinase isozyme M2	107554	1	3.2	+			
Retinoic acid receptor responder	4506425	3	3.8			+	T2
S100 A11 (calgizzarin; MLN70)	5032057	1	2.4	+			
S100 A9 (calgranulin B)	4506773	3	4.5				
Thymosin β-4	136579	1	2.8				
Transcobalamin I	4507407	2	2.6		+	+	
Transferrin[f]	4557871	6	4.2		+	+	
Tumor-assoc. antigen 90K (Mac2 BP)	5031863	4	3.5	+	+	+	
Tumor necrosis factor receptor (55 kDa)	4507575	1	3.3			+	T1
Urine protein 1	256397	1	2.6				

[a] Number of non-redundant peptides identified for each protein.
[b] Maximum cross-correlation (X_{corr}) value, using SEQUEST software, obtained for any of the peptides identified for a single protein.
[c] Levels of these proteins have been reported to be altered in serum or tumor samples from breast cancer patients.
[d] Proteins identified as potentially glycosylated in the SwissProt database.
[e] Proteins identified as probable membrane proteins in the SwissProt database. T1 and T2 indicate types I and II membrane proteins, respectively. GPI indicates that CD14 contains a covalently linked glycosylphosphatidylinositol group that anchors the protein to the membrane. 5TM indicates that CD34 belongs to a family of five transmembrane domain proteins.
[f] Proteins identified by LC-tandem MS and/or from trypsin in-gel digestion. All other proteins identified by LC-tandem MS only.

exchange by dialysis) or were fractionated using ion-exchange chromatography prior to MS analysis. The tryptic peptides from all samples were separated by cLC and the eluant directly analyzed using electrospray ionization and tandem MS.

Results of these latter analyses as well as the proteins identified using in-gel trypsin digestion are summarized in Table 1. The SEQUEST algorithm [18] was run on each of the data sets using a modified version of the human.fasta database from the National Center for Biotechnology Information. A peptide was considered to be a match based on conservative criteria developed by Yates and coworkers [19, 20]. As an additional check on the accuracy of these criteria, measured normalized elution times of peptides were compared with predicted normalized elution times calculated by an artificial neural network model developed at the Pacific Northwest National Laboratory [21]. Included in this analysis of the elution times were the 423 unique peptides that were accepted by the criteria developed in the Yates' laboratory, as well as 19 unique peptides that passed all of the criteria except the manual evaluation of the spectra. There were 11 (10 accepted, 1 rejected) peptides that eluted before a measured normalized elution time of 0.2 that were not well predicted by the model, possibly due to differences in void volume between LC separations. Although these 11 peptides were subsequently removed from the analysis, since all 10 of the accepted peptides came from proteins identified by at least 3 peptides, the final list of NAF proteins that were accepted based on the remaining 412 peptides was not affected by this exclusion.

A graph of the predicted versus measured elution times is shown in Figure 2. The predicted normalized elution times for all peptides varied by an average of 3.5% from the measured normalized elution time, indicating that the model was reliable in predicting elution times. Arrows indicate peptides that were visually identified as outliers from the main pool of peptides (Figure 2). Six of 18 (33%) of the peptides that passed all criteria but manual evaluation of the tandem MS spectra were identified as outliers. In contrast, none of the 412 peptides that were accepted could be clearly identified as outliers. Therefore, none of the proteins identified by the Yates' criteria were rejected based on the evaluation of peptide elution times. However, the evaluation of the elution times did serve to confirm that the stringency of criteria used to evaluate the tandem MS spectra was sufficient to identify proteins with a high degree of confidence. Therefore, the 64

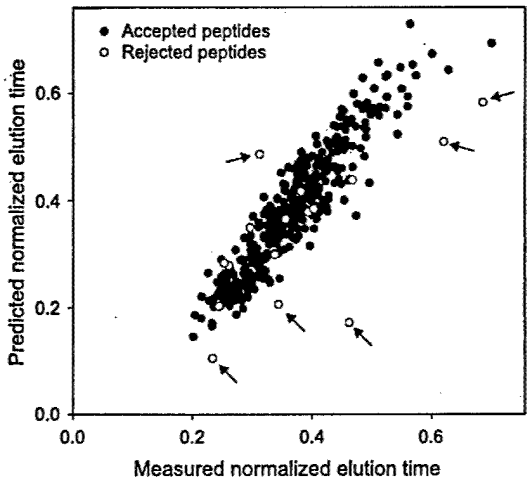


Figure 2. Evaluation of peptide elution time from the liquid chromatography column. The elution time predicted from an artificial neural network is compared to the measured elution time for the same peptide. In cases where a peptide was identified in multiple LCQ-MS analyses, the mean value of the measured elution times was used. All peptides passed various criteria based on SEQUEST parameters, as described in Materials and methods. Rejected peptides were subsequently eliminated from the pool of accepted peptides based on manual evaluation of individual tandem MS spectra. Arrows indicate the six rejected peptides that appear to be outliers from the core of the graphed data.

NAF proteins shown in Table 1 are almost certainly correctly identified by our analysis; although this list is likely biased towards moderate or high abundance proteins, which typically give better quality tandem mass spectra.

Discussion

A total of 64 proteins were identified in the NAF samples (Table 1). Of these, levels of 15 proteins (23%) have been reported to be altered in serum or tumor tissue from women with breast cancer. Cathepsin D and osteopontin have been reported to be increased in serum and tumor tissue from breast cancer patients and appear to have significant prognostic value [24–27]. Levels of other proteins may also be useful in the analysis of breast cancer but have not been as extensively studied. These proteins include α_1-antichymotrypsin, whose levels are significantly increased (23%) in plasma from breast cancer patients, but not in patients with gastric cancer or malignant melanoma [28]. In the same study, levels of α_1-

antitrypsin were also reported to be elevated in the plasma of patients with several types of cancer, with the greatest increase (56%) in breast cancer patients [28]. Low apolipoprotein D levels in breast tumors have been associated with reduced survival while elevated levels of this protein have been observed in cyst fluid of women with gross cystic disease of the breast, a condition associated with increased risk of breast cancer [29, 30]. Serum levels of ceruloplasmin have been reported to be increased in patients with breast cancer, including an approximately 2-fold increase in non-invasive breast cancer [31–33]. Post-surgery levels of ceruloplasmin also may have predictive value for recurrence of breast cancer. Normal breast epithelial cells immunostained negative for clusterin but this protein was detectable in approximately 50% of atypical hyperplasias, intraductal and invasive carcinomas [34]. Tumor levels of clusterin mRNA progressively increased with tumor size and with advancing stage of breast cancer [34]. Fibrinogen has been reported to be significantly increased (18%) in plasma from breast cancer patients, with levels of fibrinogen being the highest in women with the largest tumors [35]. Down-regulation of gelsolin in breast tumors has been associated with poor prognosis [36]. α_2-Glycoprotein (Zn) is a secreted protein reported to be increased in serum in individuals with several cancer types, including those with breast cancer [37]. Serum levels of α-lactalbumin were elevated in 62 of 97 (64%) of patients with breast cancer with mean levels 2-fold greater than the control group [38]. Levels of α-lactalbumin were greatest in sera from advance (stage IV) breast cancer patients but generally fell within the normal range in 25 samples from patients with a variety of other cancers [38]. Prolactin-induced protein is increased in serum from breast cancer patients and may be predictive of relapse in metastatic disease [39]. Levels of prolactin-inducible protein in breast tumors may also have prognostic value [39, 40]. Plasma levels of a dimeric form of pyruvate kinase M2 can discriminate controls from advanced stage breast cancer patients (specificity 85%; positive predictive value 81%) and appear to be a good marker for assessing the patient response to chemotherapy [41]. S100 A11 expression has been reported to be elevated in breast tumors [42]. The tumor-associated antigen (90K) was originally identified in conditioned medium from breast cancer cells [43, 44]. This protein has subsequently been reported to be elevated in the serum of breast cancer patients and is associated with poor prognosis [45, 46].

Breast ductal and lobular cells have been reported to have a residual secretory function in non-lactating women [7]. NAF is derived from the apocrine and merocrine gland-like surface of the breast lobular-ductal system, which secretes many of the proteins found in milk, including albumin, complement factors and immunoglobins [47]. Therefore, it is not surprising that, of the NAF proteins we identified, 35 (55%) have been reported to be in present in milk (Table 1). Based on data obtained from the Swiss-Prot database, approximately 69% of the identified proteins are potentially glycosylated (Table 1). Therefore, similar to serum, NAF appears to be enriched with glycosylated proteins. Although these proteins are glycosylated, we were able to identify them by detecting non-glycosylated peptides. Since most proteins are only glycosylated on a few amino acids, once the protein is denatured, the majority of the tryptic cut sites are accessible and numerous peptides from non-glycosylated regions are available for MS analysis. Ten of the glycosylated proteins (16%) were identified in the SwissProt database as likely transmembrane or glycosylphosphatidylinositol (GPI)-anchored membrane proteins (Table 1). The presence of membrane proteins such as tumor necrosis factor receptor in NAF may be the result of proteolytic release by 'sheddases' located on the cell surface. Alternatively, at least two of the potential membrane proteins (i.e., polymeric immunoglobin receptor and prostatin) also can be directly secreted and therefore may not have been shed.

NAF samples typically contain a small number of cells and these cells would be expected to be lysed when the NAF samples were frozen and thawed [48, 49]. We were unable to detect ribosomal proteins, histones, tubulin and myosin, which we have found in abundance in proteomic analyses of lysed cells. Therefore it is unlikely that the proteins we detect resulted from cell lysis.

Overall, our results demonstrate that NAF is a highly concentrated source of protein. Using gel-free proteomic analysis, we were able to identify 64 NAF proteins. Most likely, the proteins identified in this study are moderately abundant to abundant proteins in NAF. The identification of a number of potential biomarkers in this pool of proteins is consistent with the concept that NAF is a concentrated source of biomarkers for breast cancer. Since many of these proteins are increased in the serum of women with cancer, it seems likely that at least some of them will be increased in NAF from women with breast cancer. An important difference between analyses of the two fluids, how-

ever, will be the high level of confidence that a protein identified in NAF originated from breast tissue. We have recently developed a microarray-based ELISA [10] that allows for the simultaneous, quantitative analysis of hundreds of proteins even in small-volume samples such as NAF. Therefore, in combination with the data provided in this report, it is now possible to start performing high-throughput analyses to critically evaluate levels of proteins in NAF to determine if they are suitable as markers for breast cancer.

Acknowledgements

We gratefully acknowledge the discussions and help of Joshua Adkins and David Wunschel. This work was supported by the Biotechnology section of Core-Technology, Battelle Memorial Institute and by the Department of the Army's Breast Cancer Research Program, award no. DAMD17-00-1-0132. The content of this work does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

1. Howe HL, Wingo PA, Thun MJ, Ries LA, Rosenberg HM, Feigal EG, Edwards BK: Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst 93: 824–842, 2001
2. Leitch AM: Controversies in breast cancer screening. Cancer 76: 2064–2069, 1995
3. Ciatto S, Rosselli, Del Turco, Zappa M: The detectability of breast cancer by screening mammography. Br J Cancer 71: 337–339, 1995
4. Anonymous. Breast Cancer Progress Review Group Charting the Course: Priorities for Breast Cancer Research, 1998
5. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, Gale A: Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 94: 369–375, 2002
6. Ferrini R, Mannino E, Ramsdell E, Hill L: Screening mammography for breast cancer: American College of Preventive Medicine practice policy statement. Am J Prev Med 12: 340–341, 1996
7. Wrensch MR, Petrakis NL, Gruenke LD, Ernster VL, Miike R, King EB, Hauck WW: Factors associated with obtaining nipple aspirate fluid: analysis of 1428 women and literature review. Breast Cancer Res Treat 15: 39–51, 1990
8. Shao ZM, Nguyen M: Nipple aspiration in diagnosis of breast cancer. Semin Surg Oncol 20: 175–180, 2001
9. Sauter ER, Ross E, Daly M, Klein-Szanto A, Engstrom PF, Sorling A, Malick J, Ehya H: Nipple aspirate fluid: a promising non-invasive method to identify cellular markers of breast cancer risk. Br J Cancer 76: 494–501, 1997
10. Woodbury RL, Varnum SM, Zangar RC: Elevated HGF levels in sera from breast cancer patients detected using a protein microarray ELISA. J Proteome Res 1: 233–237, 2002
11. Malatesta M, Mannello F, Bianchi G, Sebastiani M, Gazzanelli G: Biochemical and ultrastructural features of human milk and nipple aspirate fluids. J Clin Lab Anal 14: 330–335, 2000
12. Klein P, Glaser E, Grogan L, Keane M, Lipkowitz S, Soballe P, Brooks L, Jenkins J, Steinberg SM, DeMarini DM, Kirsch I: Biomarker assays in nipple aspirate fluid. Breast J 7: 378–387, 2001
13. Covington C, Mitchell-Gieleghem A, Lawson D, Eto I, Grubbs C: Presence of carotenoid, an anticarcinogenic marker, in nipple aspirates postlactation. Adv Exp Med Biol 501: 143–152, 2001
14. Nantais-Smith LM, Covington CY, Nordstrom-Klee BA, Grubbs CJ, Eto I, Lawson DM, Pieper BA, Northouse LL: Differences in plasma and nipple aspirate carotenoid by lactation status. Nurs Res 50: 172–177, 2001
15. Harris JR: Diseases of the Breast. Lippincott, Williams & Wilkins, Philadelphia, 2000
16. Smith PK: Measurement of protein using bicinchoninic acid. Anal Biochem 150: 76–85, 1985
17. Zangar RC, Kimzey AL, Okita JR, Wunschel DS, Edwards RJ, Kim H, Okita RT: Cytochrome P450 3A conjugation to ubiquitin in a process distinct from classical ubiquitination pathway. Mol Pharmacol 61: 892–904, 2002
18. Eng JK, McCormack AL, Yates III JR: An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5: 976–989, 1994
19. Washburn MP, Wolters D, Yates III JR: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19: 242–247, 2001
20. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates III JR: Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 17: 676–682, 1999
21. Petritis K, Kangas LJ, Ferguson PL, Anderson GA, Pasa-Tolic L, Lipton MS, Auberry KJ, Strittmatter E, Shen Y, Zhao R, Smith RD: Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. Anal Chem 75: 1039–1048, 2003
22. Mannello F, Malatesta M, Sebastiani M, Battistelli S, Gazzanelli G: Molecular forms and ultrastructural localization of prostate-specific antigen in nipple aspirate fluids. Clin Chem 45: 2263–2266, 1999
23. Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD, Springer DL, Pounds JG: Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. Mol Cell Proteomics 1: 947–955, 2002
24. Furger KA, Menon RK, Tuck AB, Bramwelll VH, Chambers AF: The functional and clinical roles of osteopontin in cancer and metastasis. Curr Mol Med 1: 621–632, 2001
25. Tuck AB, Chambers AF: The role of osteopontin in breast cancer: clinical and experimental studies. J Mammary Gland Biol Neoplasia 6: 419–429, 2001
26. Rochefort H, Garcia M, Glondu M, Laurent V, Liaudet E, Rey JM, Roger P: Cathepsin D in breast cancer: mechanisms and clinical applications, a 1999 overview. Clin Chim Acta 291: 157–170, 2000
27. Mirza AN, Mirza NQ, Vlastos G, Singletary SE: Prognostic factors in node-negative breast cancer: a review of studies with

sample size more than 200 and follow-up more than 5 years. Ann Surg 235: 10–26, 2002
28. Wojtukiewicz MZ, Rucinska M, Kloczko J, Dib A, Galar M: Profiles of plasma serpins in patients with advanced malignant melanoma, gastric cancer and breast cancer. Haemostasis 28: 7–13, 1998
29. Diez-Itza I, Vizoso F, Merino AM, Sanchez LM, Tolivia J, Fernandez J, Ruibal A, Lopez-Otin C: Expression and prognostic significance of apolipoprotein D in breast cancer. Am J Pathol 144: 310–320, 1994
30. Rassart E, Bedirian A, Do, Carmo S, Guinard O, Sirois J, Terrisse L, Milne R: Apolipoprotein D. Biochim Biophys Acta 1482: 185–198, 2000
31. Vaidya SM, Kamalakar PL: Copper and ceruloplasmin levels in serum of women with breast cancer. Indian J Med Sci 52: 184–187, 1998
32. Schapira DV, Schapira M: Use of ceruloplasmin levels to monitor response to therapy and predict recurrence of breast cancer. Breast Cancer Res Treat 3: 221–224, 1983
33. Yenisey C, Fadiloglu M, Onvural B: Serum copper and ceruloplasmin concentrations in patients with primary breast cancer. Biochem Soc Trans 24: 321S, 1996
34. Redondo M, Villar E, Torres-Munoz J, Tellez T, Morell M, Petito CK: Overexpression of clusterin in human breast carcinoma. Am J Pathol 157: 393–399, 2000
35. Blann AD, Gurney D, Wadley M, Bareford D, Stonelake P, Lip GY: Increased soluble P-selectin in patients with haematological and breast cancer: a comparison with fibrinogen, plasminogen activator inhibitor and von Willebrand factor. Blood Coagül Fibrinolysis 12: 43–50, 2001
36. Winston JS, Asch HL, Zhang PJ, Edge SB, Hyland A, Asch BB: Downregulation of gelsolin correlates with the progression to breast carcinoma. Breast Cancer Res Treat 65: 11–21, 2001
37. Bundred NJ, Scott WN, Davies SJ, Miller WR, Mansel RE: Zinc alpha-2 glycoprotein levels in serum and breast fluids: a potential marker of apocrine activity. Eur J Cancer 27: 549–552, 1991
38. Thean ET, Toh BH: Serum human alpha-lactalbumin as a marker for breast cancer. Br J Cancer 61: 773–775, 1990
39. Farghaly SA: Tumor markers in gynecologic cancer. Gynecol Obstet Invest 34: 65–72, 1992
40. Pagani A, Sapino A, Eusebi V, Bergnolo P, Bussolati G: PIP/GCDFP-15 gene expression and apocrine differentiation in carcinomas of the breast. Virchows Arch 425: 459–465, 1994
41. Luftner D, Mesterharm J, Akrivakis C, Geppert R, Petrides PE, Wernecke KD, Possinger K: Tumor type M2 pyruvate kinase expression in advanced breast cancer. Anticancer Res 20: 5077–5082, 2000
42. Kurt RA, Urba WJ, Schoof DD: Isolation of genes overexpressed in freshly isolated breast cancer specimens. Breast Cancer Res Treat 59: 41–48, 2000
43. Koths K, Taylor E, Halenbeck R, Casipit C, Wang A: Cloning and characterization of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain. J Biol Chem 268: 14245–14249, 1993
44. Ullrich A, Sures I, D'Egidio M, Jallal B, Powell TJ, Herbst R, Dreps A, Azam M, Rubinstein M, Natoli C: The secreted tumor-associated antigen 90K is a potent immune stimulator. J Biol Chem 269: 18401–18407, 1994
45. Iacobelli S, Arno E, Sismondi P, Natoli C, Gentiloni N, Scambia G, Giai M, Cortese P, Panici PB, Mancuso S: Measurement of a breast cancer associated antigen detected by monoclonal antibody SP-2 in sera of cancer patients. Breast Cancer Res Treat 11: 19–30, 1988
46. Iacobelli S, Sismondi P, Giai M, D'Egidio M, Tinari N, Amatetti C, Di Stefano P, Natoli C: Prognostic value of a novel circulating serum 90K antigen in breast cancer. Br J Cancer 69: 172–176, 1994
47. Turner MD, Rennison ME, Handel SE, Wilde CJ, Burgoyne RD: Proteins are secreted by both constitutive and regulated secretory pathways in lactating mouse mammary epithelial cells. J Cell Biol 117: 269–278, 1992
48. King EB, Chew KL, Petrakis NL, Ernster VL: Nipple Aspirate Cytology for the Study of Breast-Cancer Precursors. J Nat Cancer Inst 71: 1115–1121, 1983
49. Dooley WC, Ljung BM, Veronesi U, Cazzaniga M, Elledge RM, O'Shaughnessy JA, Kuerer HM, Hung DT, Khan SA, Phillips RF, Ganz PA, Euhus DM, Esserman LJ, Haffty BG, King BL, Kelley MC, Anderson MM, Schmit PJ, Clark RR, Kass FC, Anderson BO, Troyan SL, Arias RD, Quiring JN, Love SM, Page DL, King EB: Ductal lavage for detection of cellular atypia in women at high risk for breast cancer. J Nat Cancer Inst 93: 1624–1632, 2001

Address for offprints and correspondence: Richard C. Zangar, Pacific Northwest National Laboratory, PO Box 999, P7-56, Richland, WA 99352, USA; *Tel.:* +1-509-376-8596; *Fax:* +1-509-376-6767; *E-mail:* richard.zangar@pnl.gov