

AD _____

Award Number: DAMD17-03-1-0429

TITLE: Proteomics Analysis to Identify Novel Circulating Breast Cancer Markers

PRINCIPAL INVESTIGATOR: Francisco J. Esteva, M.D., Ph.D.

CONTRACTING ORGANIZATION: The University of Texas M.D. Anderson Cancer Center
Houston, TX 77030

REPORT DATE: June 2005

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20060706107

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 01-06-2005		2. REPORT TYPE Annual		3. DATES COVERED (From - To) 27 May 2004 – 26 May 2005	
4. TITLE AND SUBTITLE Proteomics Analysis to Identify Novel Circulating Breast Cancer Markers				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER DAMD17-03-1-0429	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Francisco J. Esteva, M.D., Ph.D. E-mail:				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Texas M.D. Anderson Cancer Center Houston, TX 77030				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. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Twenty serum samples from healthy women and breast cancer patients at different stages were fractionated using two separate antibody columns to remove highly abundant proteins. Samples were randomized prior to fractionation and mass spectrometry testing. Each fraction was digested with trypsin and subsequently analyzed by LC-MS. Peptides were targeted based on the disease to control peak intensity ratios measured in the averages of all mass spectra in each group and t-tests of the intensity of each individual peak. A series of preprocessing steps were employed to produce an expansive list of peptides for further investigation and sequencing. These steps included spectral alignment, baseline subtraction, normalization, identifying of local maxima, further identifying "large" maxima as peaks, and looking for signs of differential expression. The antibody columns removed 20 of the most abundant proteins in serum. Using LC-MS and bioinformatics analysis we found 12 differentially expressed peaks in the Cancer vs. Healthy groups. Efforts are ongoing to identify targeted peptide ion signals using tandem matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS/MS). Serum fractionation using specific antibody columns followed by LC-MS and bioinformatics analysis is a feasible approach to peptide profiling in health women and breast cancer patients.					
15. SUBJECT TERMS Breast neoplasms, proteomics, serum, biological tumor marker					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 8	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Table of Contents

Cover.....	1
SF 298.....	2
Introduction.....	4
Body	4
Key Research Accomplishments.....	7
Reportable Outcomes.....	8
Conclusions.....	8
References.....	8
Appendices.....	N/A

Introduction

Serum protein profiling using mass spectrometry is a promising approach to identify novel circulating breast cancer markers. One of the major problems with detecting low-abundance proteins in the serum is that they are frequently masked by large, abundant proteins such as albumin and immunoglobulins among others. Therefore, serum protein fractionation is an important consideration. After fractionation, protein profiles can be detected using mass spectrometry. Surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) has been used to compare protein profiling of serum from healthy individuals and cancer patients. However, SELDI-TOF only yields mass/charge (effectively molecular weight) information and no protein identification. Alternatively, fractionated serum proteins can be analyzed after protease digestion using liquid chromatography mass spectrometry (LC-MS), and the LC-MS profiles can then be compared to develop diagnostic fingerprints using bioinformatic techniques. Differentially regulated peptides can then be identified by MS/MS, allowing verification and antibody-based diagnostics to be developed.

Body

Thirty serum samples from healthy women and breast cancer patients at different stages were fractionated using two separate antibody columns to remove highly abundant proteins. Samples were randomized prior to fractionation and mass spectrometry testing. Briefly, 20 microliters of serum were diluted and injected through a Seppro column and an Agilent column in tandem using appropriate buffers. Each fraction was digested with trypsin and subsequently analyzed by LC-MS. Rather than using bioinformatic analysis as a pattern-matching technique, peptides were targeted based on the disease to control peak intensity ratios measured in the averages of all mass spectra in each group and t-tests of the intensity of each individual peak. A series of preprocessing steps were employed to produce an expansive list of peptides for further investigation and sequencing. These steps included spectral alignment, baseline subtraction, normalization, identifying of local maxima, further identifying "large" maxima as peaks, and looking for signs of differential expression (Koomen, et. Al, 2005).

Serum samples were obtained under protocol LAB02277 (UTMDACC) with appropriate consent forms on file, aliquoted, and stored frozen at -80°C . Aliquots (20 μl) from each were separately thawed, diluted 5x in TBS (20 mM pH7.6) and injected onto the depletion columns (Agilent-6, Seppro-12) in tandem flowing at 200 μl per minute in TBS. The effluent was monitored at 280 nm and the flowthrough was collected. The affinity column system was flushed with loading buffer, regenerated with 500 mM Glycine-HCl pH2.0 in TBS and reequilibrated in TBS for the next sample injection. Pilot experiments indicated sample carryover under these conditions was essentially undetectable. The above flowthrough was acetone-precipitated by adding 6 volumes of cold (-20°C) acetone and standing at -20°C overnight. The liquid was carefully decanted, the pellet was washed once with cold (-20°C) acetone, and the pellet air-dried for several minutes. To this 500 μg trypsin (sequencing grade, Promega) was added in 50 μl 30 mM ammonium bicarbonate and the digestion proceeded for 8 hours at 37°C , after which an additional 500 μg trypsin was added and incubated overnight. The digestion was quenched by the addition of acid, and 5 μl injected on the LCMS for profiling.

LCMS was performed using a capillary HPLC (Agilent 1100 capillary) connected to an ESI-TOF mass spectrometer using a nanoflow interface (Mariner, Applied Biosystems). The separation was performed on a 0.150 mm ID x 15 cm C18 reversed-phase column (C18- MS, Grace-Vydac) flowing at 1 $\mu\text{L}/\text{min}$. Samples were injected at 97% A (2% acetonitrile in water containing 0.01% trifluoroacetic acid), and salts flushed out for 40 minutes. Then the mass spectral acquisition was started with the gradient start, proceeding to 50% B (80% aqueous acetonitrile containing 0.01% trifluoroacetic acid) over 40 minutes, then ramping up to 90% B over 5 minutes. After flushing at 90% the column was reequilibrated in initial conditions, and two blank gradients were performed to reduce the possibility of peptide carryover into the next run. Preliminary experiments indicated this protocol was more than sufficient for this purpose.

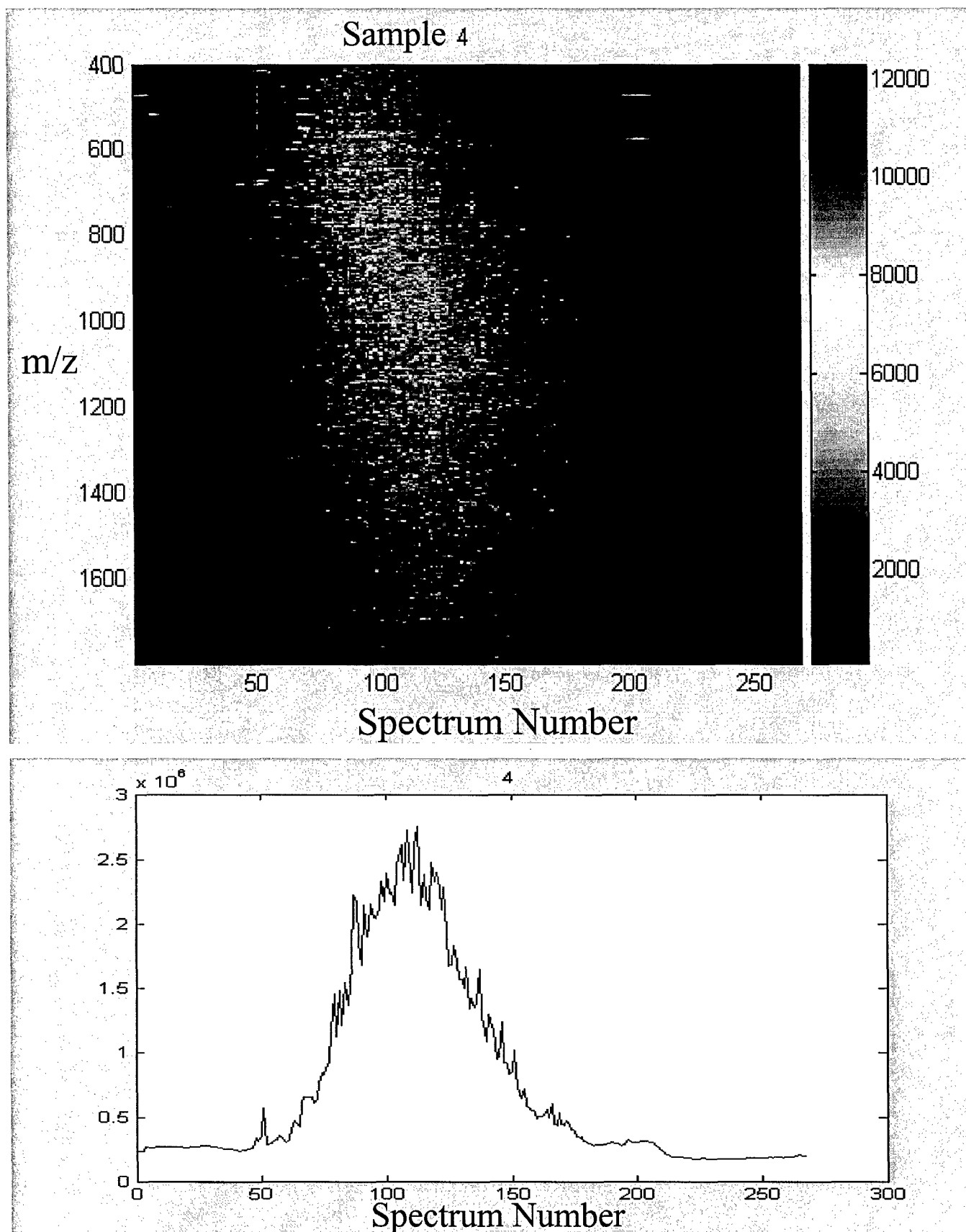


Figure 1. Heat map (upper panel) and total ion chromatogram (lower panel) of one of the samples. Heat map intensity color-code is to the right, TIC ordinate is in counts.

Mass spectra were acquired as the sum of 20 seconds of elution time per spectrum over the course of the 90 minute run, resulting in about 270 spectra per sample. A heat map of the LCMS one of the samples is shown in figure 1 (upper panel), above. The corresponding total ion chromatogram (TIC) is also shown in figure 1 (lower panel). We found there was some variation in the retention times of several major peptide signals, so we adjusted the time coordinates slightly based on apparent retention times of a number of peaks identified as originating from an abundant protein, complement 3. We then calculated the offsets in various regions of the chromatogram, and performed a piece-wise adjustment to the apparent retention times for each run. An example of the adjustment is illustrated in figure 2, below.

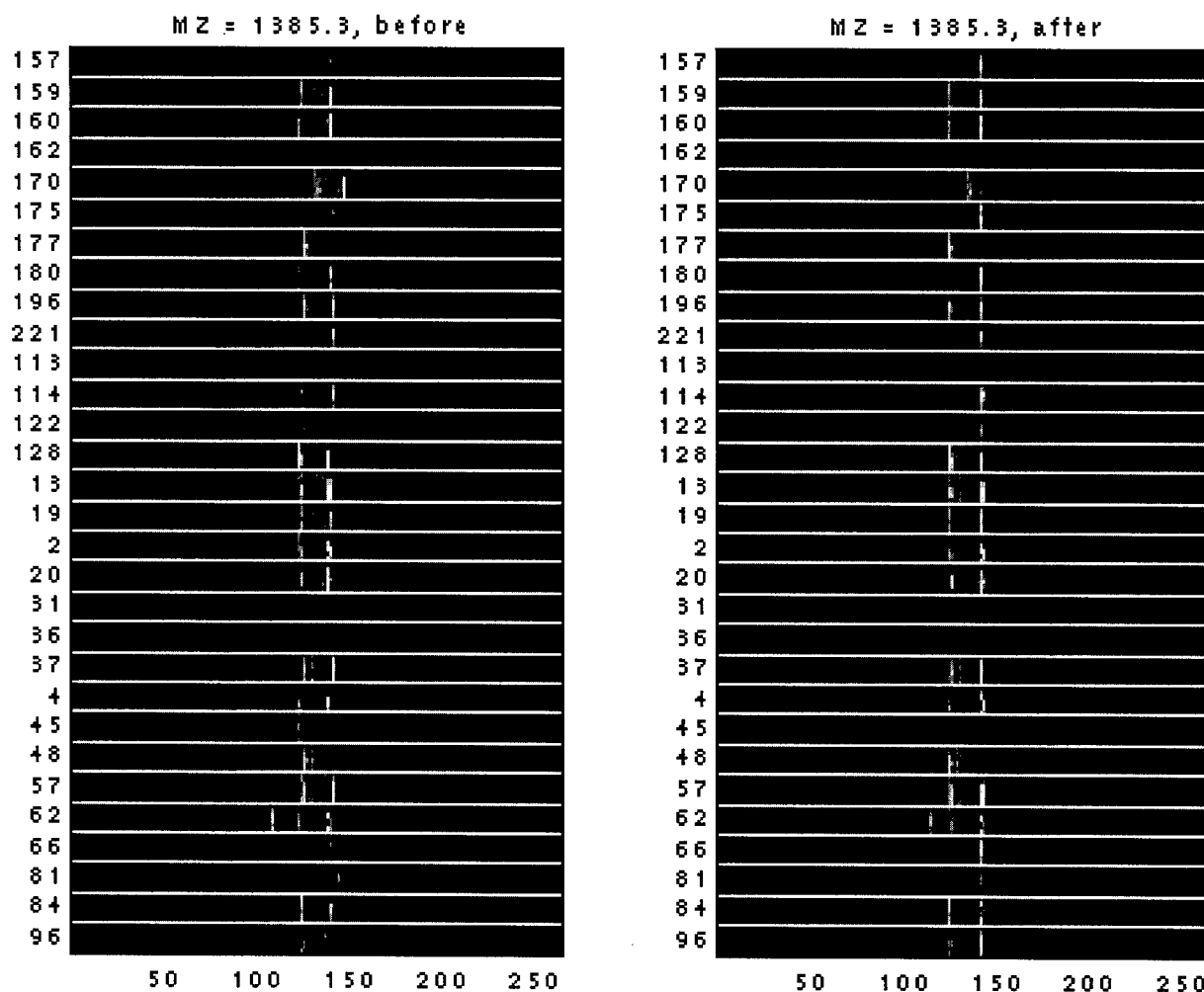


Figure 2. Adjusting the retention time in the neighborhood of the 1385.3 peak. Unadjusted data for mass 1385.3 across the sample set is on the left, the right panel shows the result of the time adjustment.

Our early analysis of these data generated lists of peaks that appeared to be up- or down- regulated based on their t-scores. One of them, expected to have a molecular weight of 1339.7 was found in sample number 48. A portion of this digest was then fractionated and analyzed by LC-MALDI-MS/MS (Dionex-LCPackings HPLC with Probot plate spotting robot, Applied Biosystems 4700 Proteomics Analyzer). Approximately 50 proteins were identified in this experiment with reasonable confidence levels. Of these, one of the proteins found was Protein S. This protein was identified on the basis of a single peptide match, which had the correct MH+

(1340.8) corresponding to the Mr of 1339.7. The match score using the search-engine Mascot was 69, normally a very good score. The spectrum match generated by Mascot is shown in figure 3.

{MATRIX} *{SCIENCE}* Mascot Search Results

Peptide View

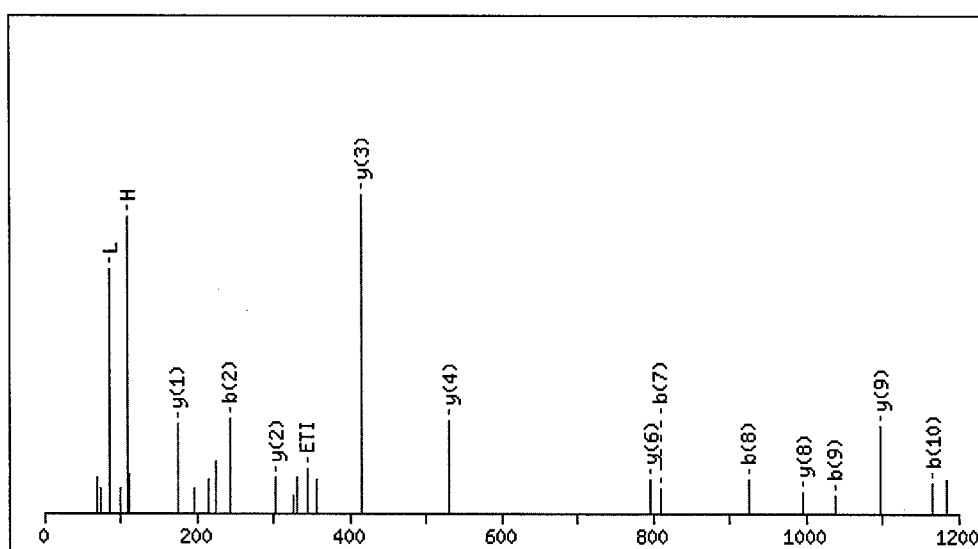
MS/MS Fragmentation of IETISHEDLQR

Found in gi|36579, preproprotein S [Homo sapiens]

Match to Query 428 (1340.80,1+) MascotID: 21866, SpectrumID: 65190,

Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da



Monoisotopic mass of neutral peptide (Mr): 1339.67

Figure 3. Centroided spectrum match output from Mascot for the target peptide at $MH^+=1340.8$. The score for this match was 69.

Further statistical analyses of the dataset subsequently revealed that the number of dysregulated peaks we found was actually no greater than the number expected by pure chance. We believe that the depletion experiment removed most of the proteins often found to be significantly dysregulated in such experiments, such as haptoglobin and serum amyloid. Apparently the next level of proteins detectable by these methods are not sufficiently perturbed to be found from the noise in this system. We are now pursuing some next-generation strategies to overcome this problem.

Key Research Accomplishments

The Seppro and Agilent antibody columns removed 12 of the most abundant proteins in serum, including albumin, IgG, Fibrinogen, Transferrin, IgA, IgM, α 1-Antitrypsin, Haptoglobin, α 1-Acid Glycoprotein, α 2-Macroglobulin and HDL (Apolipoproteins A-I and A-II). Using LC-MS and bioinformatic analysis we found 17 differentially expressed peaks in the Cancer vs. Healthy groups; 28 differentially expressed peaks in the Stage 3 vs. Healthy groups; and 36 differentially expressed peaks in the Stage 4 vs. Healthy groups. Efforts are ongoing

- to identify targeted peptide ion signals using tandem matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS/MS). One peak indicated to be up-regulated in cancer by the initial bioinformatics analysis was identified as a peptide from Protein S, but it is statistically unconvincing. Further improvements are required to find convincing biomarker candidates.

Reportable Outcomes

None

Conclusions

Serum fractionation using specific antibody columns followed by LC-MS and bioinformatic analysis may be a feasible approach to peptide profiling in healthy women and breast cancer patients. A key advantage is that detected changes can be identified by ms/ms of the target peptides. A disadvantage compared with the SELDI experiment is that each sample produces about 100 times more data per sample to process. Still, further improvements in processing analysis appear to be necessary to produce convincing candidate biomarkers.

References

Koomen JM, Zhao H, Li D, Nasser W, Hawke DH, Abbruzzese JL, Baggerly KA, Kobayashi R. Diagnostic protein discovery using liquid chromatography/mass spectrometry for proteolytic peptide targeting. Rapid Commun Mass Spectrom. 2005;19(12):1624-36.

Appendices

N/A