

Award Number: W81XWH-05-1-00298

TITLE: Proteomic Prediction of Breast Cancer Risk: A Cohort Study

PRINCIPAL INVESTIGATOR: Thomas E. Rohan, Ph.D.

CONTRACTING ORGANIZATION: Albert Einstein College of Medicine  
Bronx, NY 10461

REPORT DATE: March 2006

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.

# 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.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 01-03-2006		<b>2. REPORT TYPE</b> Annual		<b>3. DATES COVERED (From - To)</b> 1 Mar 2005 - 28 Feb 2006	
<b>4. TITLE AND SUBTITLE</b> Proteomic Prediction of Breast Cancer Risk: A Cohort Study				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> W81XWH-05-1-0298	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Thomas E. Rohan, Ph.D.  E-Mail: <a href="mailto:rohan@aecom.yu.edu">rohan@aecom.yu.edu</a>				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  Albert Einstein College of Medicine Bronx, NY 10461				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT:</b> Our objective is to develop and test proteomic methods for the prediction of breast cancer risk, an approach that has not been attempted previously. Our underlying hypothesis is that proteomic analysis of serum will identify proteins differentially expressed in women who do versus those who do not develop invasive breast cancer, and that these differences will be identifiable prior to the clinical presentation of breast cancer. Our work will be conducted in two phases, a training phase and a test phase. Both phases will be conducted as case-control studies nested in a population-based cohort of women who were members of Kaiser Permanente and has serum specimens collected between 1986 and 1992. To date, we have finalized our cohort definition, finalized the definition of cases and controls; finalized the criteria for matching controls to cases; selected the cases and controls; and initiated the pulling, testing, and aliquoting of the serum specimens. Furthermore, we have developed an extremely detailed protocol for proteomic analysis of the serum samples. Briefly, the serum sample is loaded onto an immunoaffinity column to deplete twelve abundant proteins, and the flow-through fraction is collected and subjected to reduction, alkylation, and digestion. Subsequently, the digested proteins are labeled with iTRAQ reagents and prefractionated with cation exchange chromatography. Each fraction is loaded onto the reverse phase column and followed by MALDI-TOF/TOF (4700 Proteomic Analyzer) analyses. The data collected are automatically processed, combined, and searched against human protein database. There are no research accomplishments or reportable outcomes to date (end of year 1). Previous proteomic studies of breast cancer have used lower resolution proteomic approaches in cross-sectional settings only. By applying high-resolution proteomic approaches to a prospective setting, this ongoing project may enhance our ability to identify those women at increased risk of breast cancer and intervene before they progress to cancer. Furthermore, it may provide insight into the biological processes underlying breast cancer development, for example, by identifying the protein markers underlying case-control proteomic differences and by leading to identification of the associated candidate genes that might be responsible for differences in susceptibility.					
<b>15. SUBJECT TERMS</b> Epidemiology, Cohort Study, Proteomics					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>USAMRMC</b>
U	U	U	UU	17	<b>19b. TELEPHONE NUMBER (include area code)</b>

## Table of Contents

Cover.....	
SF 298.....	2
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	13
Reportable Outcomes.....	13
Conclusions.....	13
References.....	13
Appendices.....	14

## **Introduction**

New approaches to breast cancer risk prediction are needed given the modest discriminatory accuracy of existing risk prediction models at the individual level. Our objective is to develop and test proteomic methods for the prediction of breast cancer risk, an approach that has not been attempted previously. Our underlying hypothesis is that proteomic analysis of serum will identify proteins differentially expressed in women who do versus those who do not develop invasive breast cancer, and that these differences will be identifiable prior to the clinical presentation of breast cancer. Specifically: (1) using a training set of serum specimens sampled from a population-based cohort of women who were members of Kaiser Permanente, who had their blood samples taken at a multiphasic health examination (MHC) between 1986 and 1992, and who were followed up to determine subsequent breast cancer occurrence, we plan to apply high sensitivity proteomic approaches to identify biomarkers that discriminate between women who developed invasive breast cancer within 5 years of having a serum sample collected (cases) and women who remained free of breast cancer for at least that long (controls); (2) using a validation set of serum samples collected from a separate group of cases and controls in the cohort, we plan to test the biomarkers identified using the training set. Nested case-control studies will be undertaken at the training (40 cases/40 controls) and validation stages (20 cases/20 controls). Cases will be women with no history of breast cancer at baseline who developed a subsequent incident, invasive breast cancer within 5 years of collection of a serum sample; controls (no breast cancer history) will be selected from the same cohort using risk-set sampling and matched individually to cases on calendar year of the MHC exam at which the serum sample of interest was collected, age at that MHC exam, and time since last meal. To reduce the possible impact of breast cancer heterogeneity on interpretation of the results, we will restrict inclusion to white, postmenopausal women; should the results prove informative, we plan to study other subgroups (e.g., African American, premenopausal, etc.) in the future. Data on breast cancer risk factors will come from questionnaire and physical exam data gathered at the MHC. All proteomic spectra will be analyzed by the SEQUEST program to identify the protein sequences generating all peptide ions. After biomarker ions have been identified, validation will be carried out by the same analytical procedures following only the biomarker ions of interest. The laboratory work will be performed “blinded” to case-control status. Supervised clustering algorithms such as support vector machines will be used for the statistical analysis of the training dataset to identify candidate marker patterns that best discriminate between cases and controls. The most promising markers identified in the training set will be applied to subjects in the validation set and used to classify them as cases, controls, or neither. In contrast to the training set, where case-control status will be known to the data analysts, case-control status will not be known during analysis of the validation set. Therefore, we will determine how well the method developed at the training stage discriminates between cases and controls in the validation set.

## **Body**

### ***(i) Eligibility criteria***

We have finalized eligibility for inclusion in the cohort. We had proposed initially to use a cohort recruited between 1964 and 1971. However, because serum samples from that cohort may previously have undergone thawing and re-freezing, we decided to switch to the more recent Kaiser cohort, which was recruited between 1986 and 1992. None of the samples from the latter cohort have been thawed to date. The cohort consists of white, postmenopausal women aged 55 to 80 years at the time of the blood draw. As indicated in our original proposal, we chose to restrict attention to

white women to reduce the possible impact of breast cancer heterogeneity on interpretation of the results.

***(ii) Case/control definition and selection***

The work for this project is being conducted as nested case-control studies, both at the training and at the validation stages. Cases are defined as white, postmenopausal women with no history of breast cancer at the time of recruitment. Inclusion is restricted to subjects between the ages of 55 and 80 years. We identified 68 potentially eligible cases by merging data from the multiphasic cohort/serum repository databases with data from the Kaiser Permanente tumor registry. The 60 cases ultimately included in the study will be a random sample of all eligible cases. Controls are matched 1:1 to the corresponding case. They are white, postmenopausal women with no history of breast cancer and who have not developed breast cancer by the date of diagnosis of the corresponding case. The controls were selected using risk-set sampling with replacement. They were matched to the corresponding case on age (within 1 year), date of serum collection (to within 1 month), and time since last meal (0-3 hours, 4-9 hours). Furthermore, cases and controls were matched with respect to membership of Kaiser Permanente, in the sense that as with the cases, controls were required to have been members from one year prior to serum collection and to have been a member at the time of diagnosis of the corresponding case (those whose membership lapsed for a period exceeding 3 months ceased to be eligible to be selected as a control). Although only 1 control per case will be included in the study, we selected 2 controls for each case to allow for the possibility that the serum sample for the first control could not be located or was desiccated and hence unusable. We have assigned study identification numbers to the cases and controls, and these numbers are linkable to the Kaiser IDs, which will allow extraction of corresponding covariate information from the cohort database.

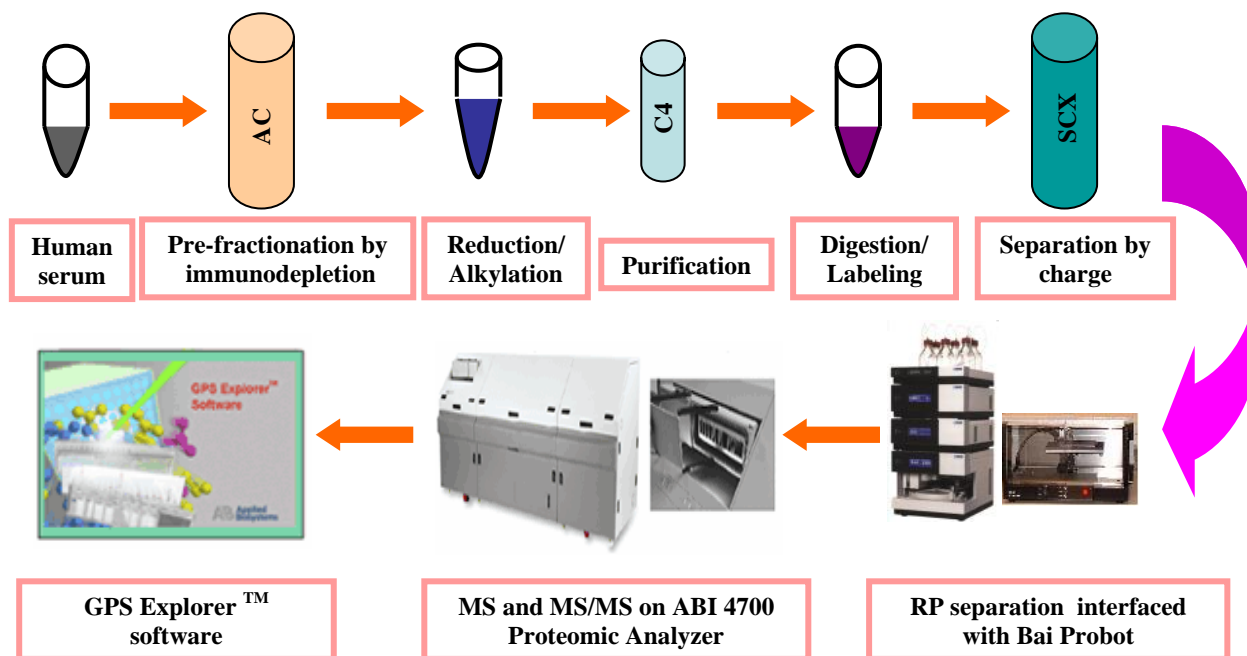
***(iii) Aliquoting of serum samples***

We have now finalized plans for the pulling, testing, and aliquoting of the specimens at the Orentreich Foundation and expect these tasks to be completed in approximately 4-6 weeks. We will be in a position to commence the proteomic analyses immediately upon receipt of the serum samples from the Orentreich Foundation, and each run, which will include an aliquot of serum from a case, its corresponding control, and the common pool [the common pool is created from all 80 subjects included in the training phase – that is, their samples are pooled and aliquots of the pool are then included in each run]), is expected to take about 1 week to complete. Hence, completion of the proteomic analysis of the training set of serum samples should require about 40 weeks, although in practice it is possible that a few weeks more will be required if some runs have to be repeated.

***(iv) Development of methods for proteomic analysis of serum samples***

In parallel with the activities described above, we have established detailed protocols and reaction conditions for each step of the proteomic analysis. Specifically, we are developing and validating an integrated, highly sensitive proteomic strategy by combining immunodepletion, extensive fractionation by multi-dimensional HPLC, and MALDI-TOF/TOF-MS, to tackle the complexity and dynamic range of the serum proteome. Here we focus on the protocol setup, which entailed use of standard human serum samples. The general workflow of our experiments is shown in Figure 1. Briefly, the serum sample is loaded onto an immunoaffinity column to deplete twelve abundant proteins, and the flow-through fraction is collected and subjected to reduction, alkylation, and digestion. Subsequently, the digested proteins are labeled with iTRAQ reagents and prefractionated with cation exchange chromatography. Each fraction is loaded onto the reverse phase column and

followed by MALDI-TOF/TOF (4700 Proteomic Analyzer) analyses. The data collected are automatically processed, combined, and searched against human protein databases.



**Figure 1.** General workflow of proteomic study of breast cancer

## Materials and Methods

**1. Reagents.** Tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) and methyl methane-thiosulfonate (MMTS) were obtained from Pierce (Rockford, IL). Triethylammonium bicarbonate (TEAB), ethanol and iTRAQ reagents were from Applied Biosystems (Forster City, CA). Sequencing Grade Modified Trypsin (TPCK treatment) was purchased from Promag (Madison, WI). Guanidium hydrochloride was purchased from USB Corporation (Cleveland, OH). Human serum sample was obtained from National Institute of Standards of Technology (Gaithersburg, MD).

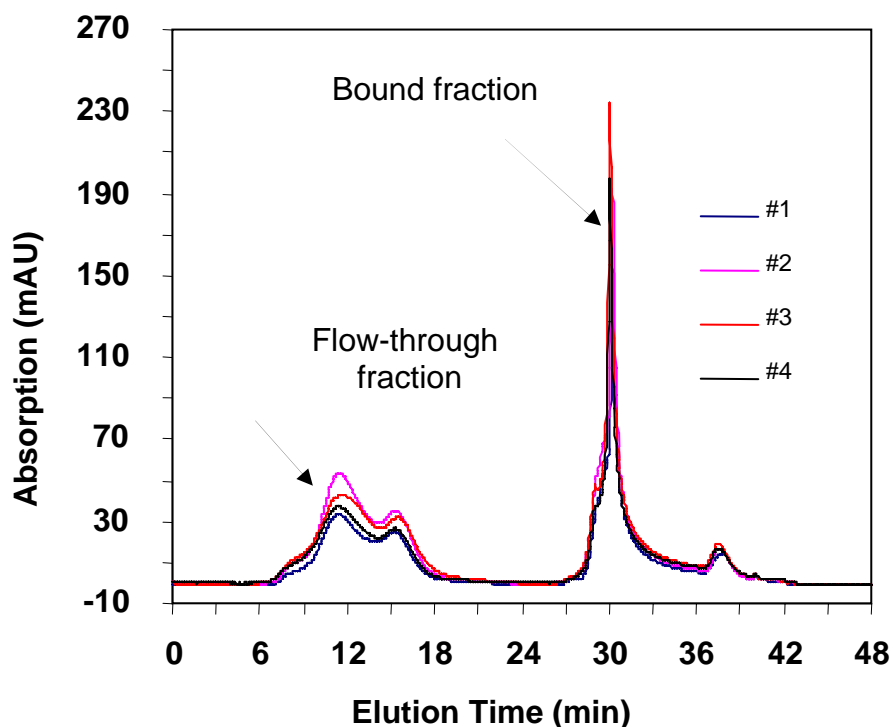
**2. Protocol.** The steps involved in our protocol are shown in the Appendix and described here.

**(a) Sample preparation** 20  $\mu$ l of serum sample were loaded onto the immunodepletion column (6.4  $\times$  63 mm) from Genway Biotech Inc (San Diego, CA) and the flow-through fraction was collected. The sample was then denatured (6M) by adding guanidium hydrochloride, reduced (2 mM) by adding Tris-(2-carboxyethyl) phosphine hydrochloride (60  $^{\circ}$ C, 1hr) and alkylated (5 mM) by adding methyl methane-thiosulfonate (rm, 1hr). The resulting solution was passed through a 2  $\times$  30 mm C<sub>4</sub> reverse phase column from Perkin Elmer (Shelton, CT) and eluted with a gradient of 0-90% buffer B (Buffer A is composed of 0.1% TFA; Buffer B is composed of 80% acetonitrile, 10% n-butanol, 0.08% TFA). The collected protein solution was concentrated by reducing volume with SpeedVac. Then the sample was 10-fold diluted in the digestion buffer (50 mM triethylammonium bicarbonate) and digested with trypsin (1:20 w/w) overnight at 37  $^{\circ}$ C. The digested sample solution volume was reduced with SpeedVac. For each digested sample, 100  $\mu$ l of labeling buffer (50 mM TEAB, 75 % ethanol) was added, after which 100  $\mu$ g of iTRAQ reagents (114, 115, 116, and 117) was added and allowed to react at room temperature for 1 hr. Residual reagent was quenched by adding 300  $\mu$ l of water and allowing excess reagent to completely hydrolyze over an additional 30 minutes, then the four labeled samples were mixed.

**(b) Off-line 2D LC coupled with MALDI-TOF/TOF analyses** The combined peptide mixture was separated by strong cation exchange (SCX) chromatography on an AKTA purifier 10 system from GE Healthcare Bio-Sciences (Piscataway, NJ) using a PolySulfoethyl A<sup>TM</sup> column (2.1 × 100 mm, 5 μm, 300Å) from Poly LC Inc (Columbia, MD). Sample was diluted in 1ml of SCX loading buffer (20% acetonitrile, 10 mM KH<sub>2</sub>PO<sub>4</sub>, pH 3, with phosphoric acid) and loaded and washed isocratically for 20 min at 0.1ml/min to remove excess reagent. Peptides were eluted with a linear gradient of 0-700 mM KCl (20% acetonitrile, 10 mM KH<sub>2</sub>PO<sub>4</sub>, pH 3) over 15 minutes at a flow rate of 0.1 ml/minute, with fractions collected at 1-minute intervals. Then peptide separation was performed on an Ultimate<sup>TM</sup> chromatography system equipped with a Bai Probot MALDI spotting device (Dionex Corp, Sunnyvale, CA). Individual SCX fractions were injected and captured onto a trap column (1× 8mm) from Michrom Bioresources Inc (Auburn, CA) and then eluted onto an RP – C<sub>18</sub> capillary column (300 μm ×150 mm) purchased from Dionex Corp (Sunnyvale, CA) with a gradient of buffer B (buffer A, 0.1% TFA, 5% H<sub>2</sub>O; buffer B, 0.1% TFA, 95% acetonitrile, 5% H<sub>2</sub>O). Column effluent was mixed automatically in a 2:1 ratio with MALDI matrix (saturated cyano-4-hydroxycinnamic acid) with a probot MALDI spotting device, spotted directly on MALDI plate. MALDI plates were analyzed on an ABI 4700 Proteomic Analyzer from Applied Biosystems (Framingham, MA).

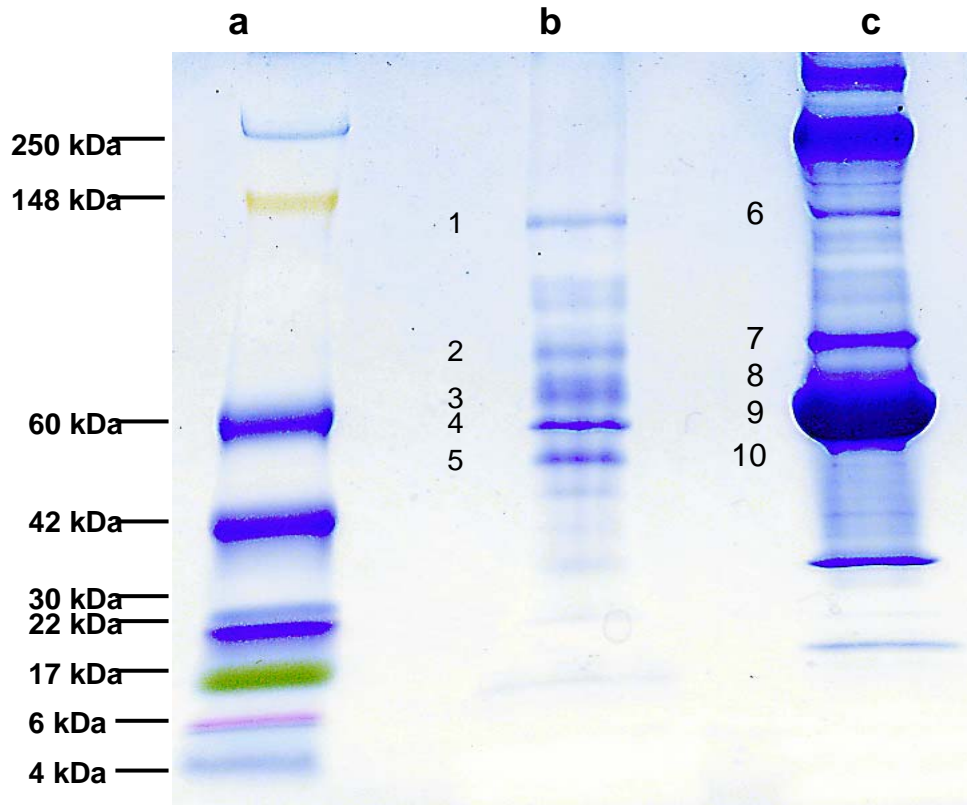
### 3. Preliminary Results

Figure 2 (below) shows the typical chromatograms of four runs obtained in the immunodepletion of standard human sera. The left peaks are the flow-through fraction and the right peaks are the bound fraction. The results are reproducible in terms of peak shape and elution time/volume. The flow-through fraction was collected, reduced, and alkylated, then passed through C<sub>4</sub> reverse phase column.

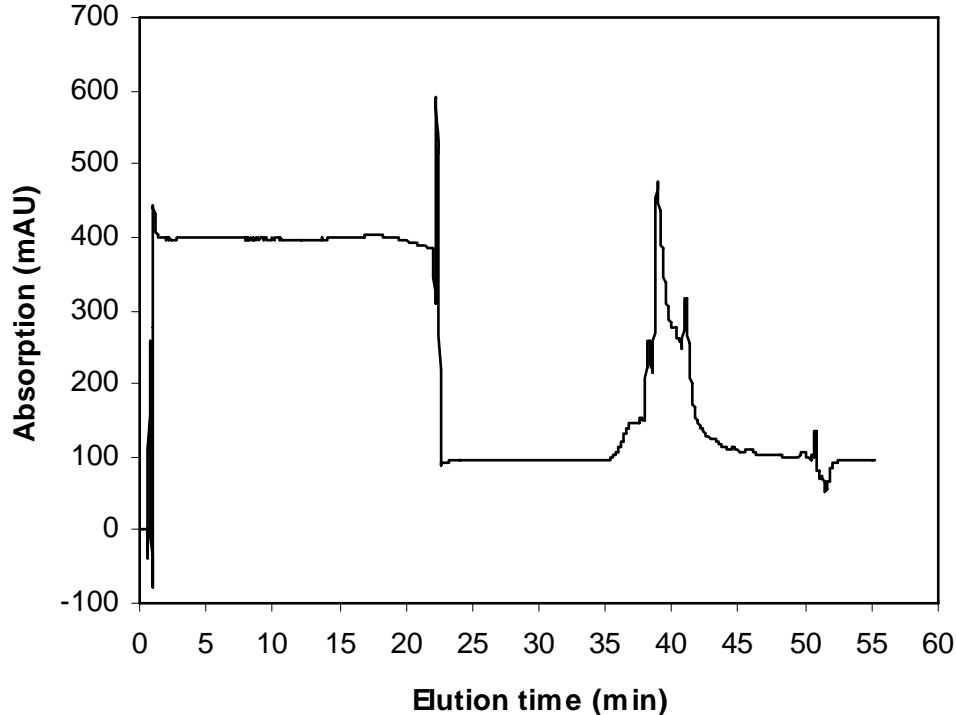


**Figure 2.** Typical chromatograms of immunodepletion of human serum. Left peaks are the flow-through fraction and right peaks are the bound fraction stripped off from immunodepletion column with a gradient of glycine.

Figure 3 (below) shows a typical chromatogram of the protein sample obtained in the sample purification with C<sub>4</sub> column. Subsequently, the sample was checked with 1D gel (Figure 4 (below)). Then in-gel digestion coupled with mass spectrometry analysis of 10 bands in Figure 4 was performed and the results demonstrated that the immunodepletion process had removed 12 abundant proteins completely.

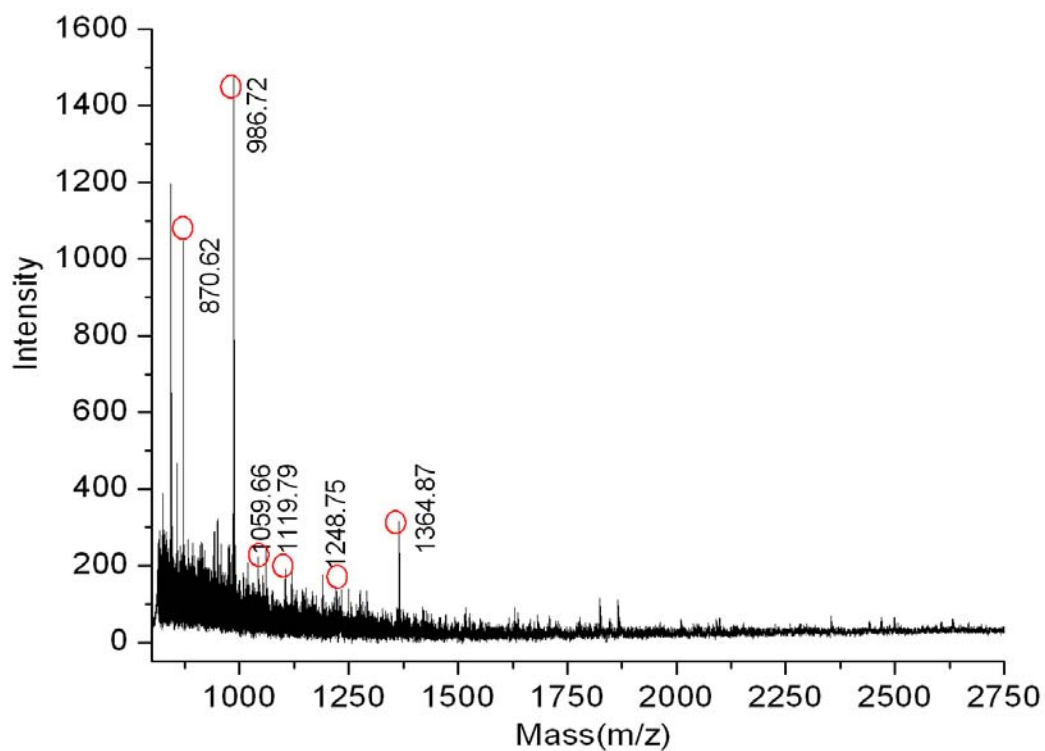


**Figure 3.** 1D 4-20% gradient SDS-PAGE of human serum sample. Lane a: Molecular weight markers. Lane b: Human serum after depleted with Seppro™ chicken antibodies (IgY) column (Genway Biotech). Lane c: Crude human serum. The in-gel digest and mass spectrometry analysis results are as follows: Band 1, Human factor H ; Band 2,  $\alpha$ 1-B-glycoprotein precursor; band 3, Unnamed proteins; band 4, Vitamin D-binding protein precursor; band 5 Apolipoprotein H; band 6, Ceruloplasmin; band 7, Transferrin; Band 8, 9, and 10, HSA.

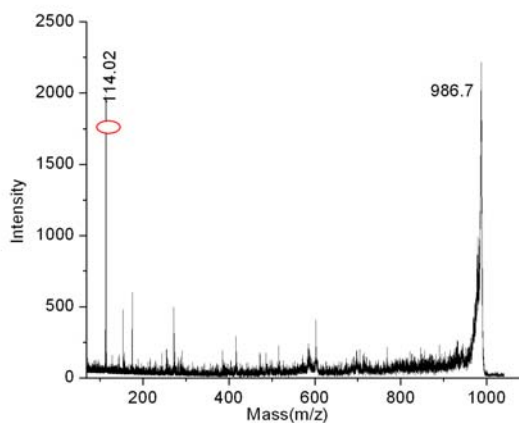
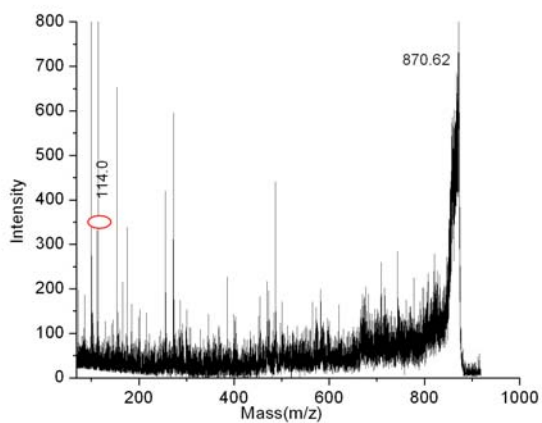


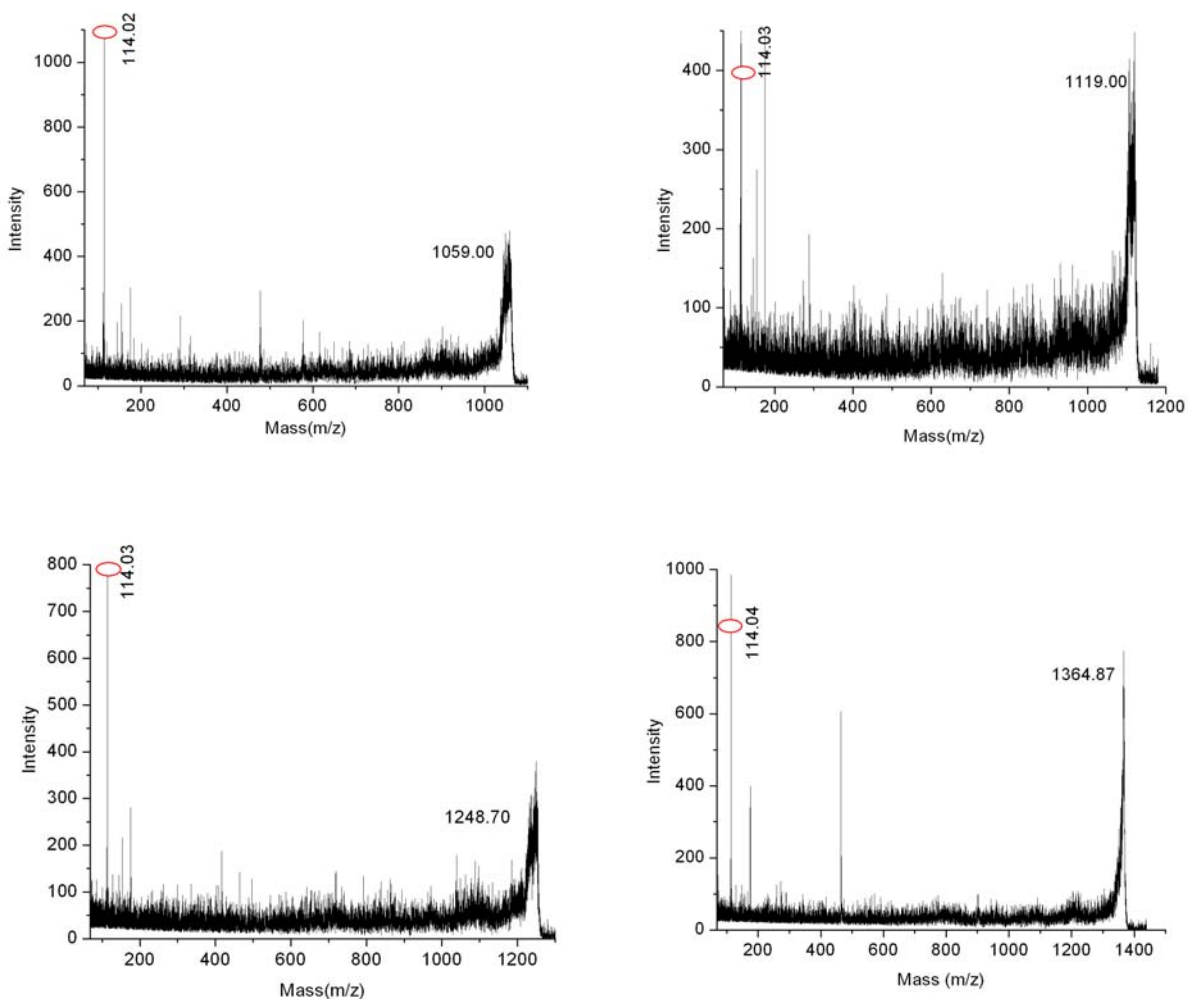
**Figure 4.** Typical chromatogram of C<sub>4</sub> purification step. Left big hump is the elution peak of 6M guanidium hydrochloride and right peaks are the purified protein peaks.

Figure 5 (below) shows a typical mass spectrum of the serum proteins after immunodepletion, digestion and iTRAQ reagent labeling. iTRAQ reagents from Applied Biosystems are amine-specific and a multiplexed set of 4 isobaric reagents and they yield labeled peptides with identical m/z in single MS mode, but produce strong, diagnostic, low-mass MS/MS signature ions allowing for quantification of up to four different samples simultaneously. Note that in Figure 5, a quarter of the starting serum sample (4  $\mu$ l) was used to obtain the spectrum and the digested labeled peptides were subjected to a micro-chromatography separation (C<sub>18</sub> ZipTip) before mass spectrometry analysis. The signals in the spectrum are not dense and strong because of signal suppression that can be partially restored through extensive LC separation. It should be pointed out that even the precursor signals in single MS mode (Figure 5) are not strong. However the signature ions (114 ions) (Figure 6) in the MS/MS mode have very strong intensity. We believe that strong signals will result if the amount of sample analyzed is increased.



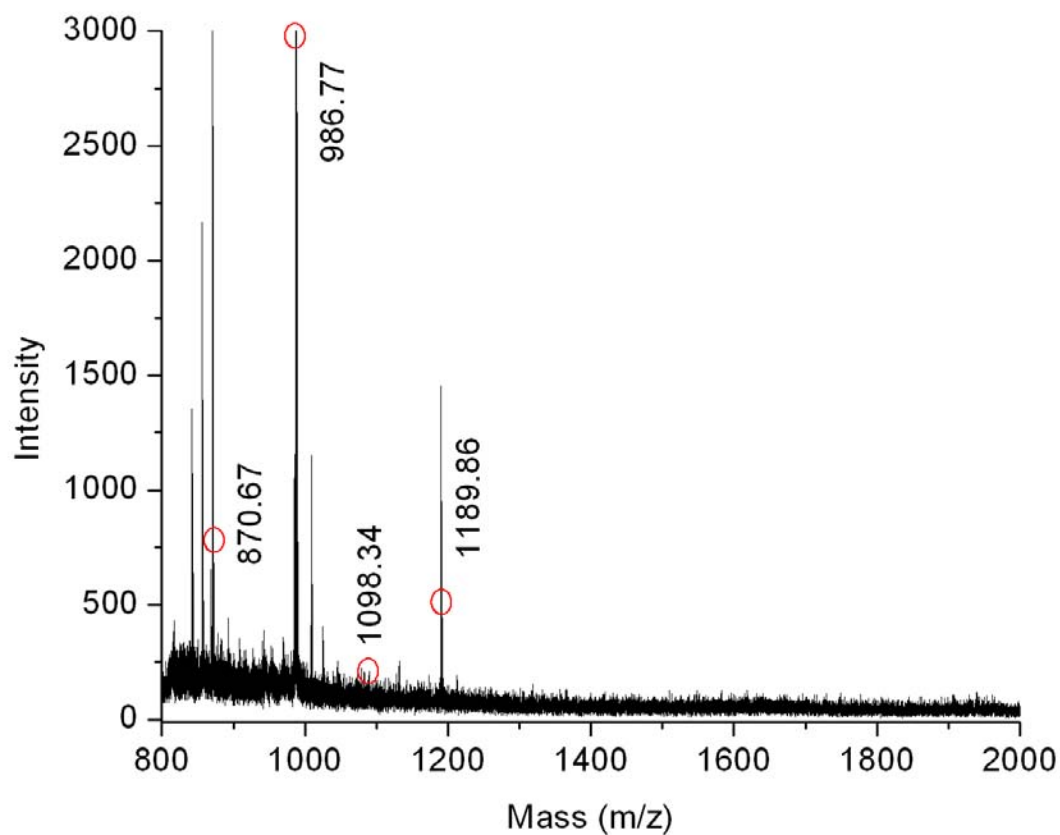
**Figure 5.** Typical mass spectrum of human serum after depletion, reduction, alkylation, digestion, and labeling with the 114 iTRAQ Reagent. The highlighted peaks were chosen as precursors to get MS/MS signature ion signals.



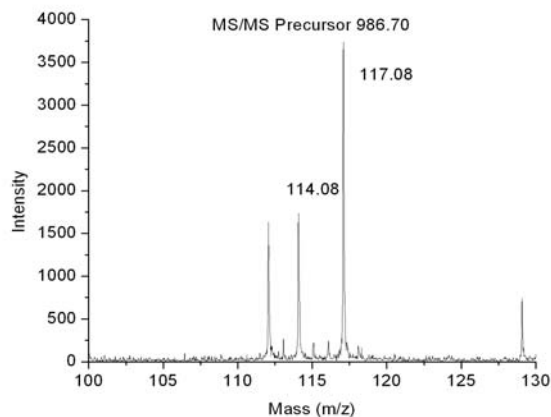
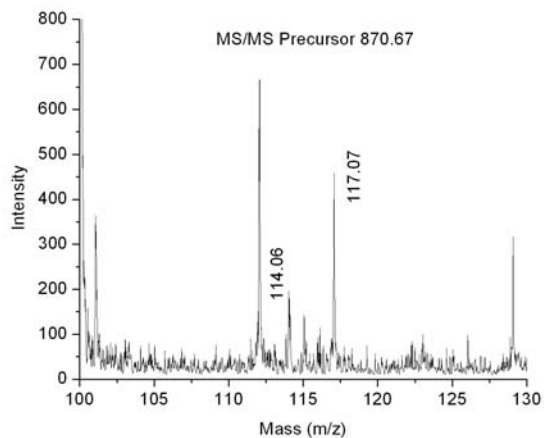


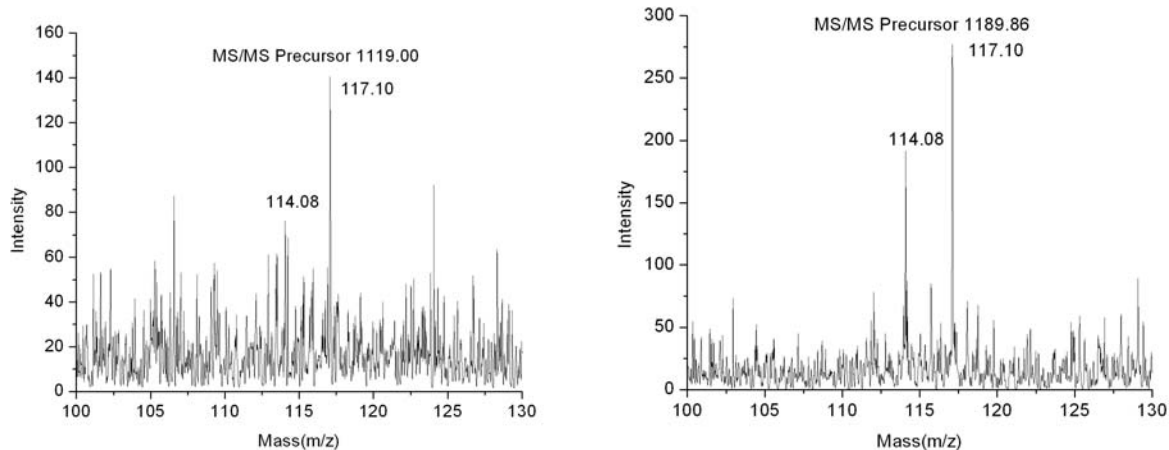
**Figure 6.** Signature ions (114) of MS/MS spectra corresponding to peptide ions 870.62, 986.7, 1059.00, 1119.00, 1248.70, 1364.87.

Our goal is to identify proteins expressed differentially in women who develop invasive breast cancer and women who remain free of breast cancer. Thus the quantitative method (iTRAQ chemistry) involving parallel workflows must be accurate and precise in order to identify the differentially expressed proteins. For simplicity, two serum samples were labeled with iTRAQ reagents 114 and 117, mixed in ratios 1:2 and quantified by comparing the intensities of the signature ions (114 and 117) (Figures 7 and 8 (below)). As shown in Figure 8, the relative ratio is accurate compared to the expected values (1:2).



**Figure 7.** Typical mass spectrum of human serum after depletion, reduction, alkylation, digestion, and labeling with the 114 and 117 iTRAQ Reagents. The highlighted peaks are chosen as precursors to get MS/MS signature ions (114 and 117).





**Figure 8.** Signature ions (114 and 117) of MS/MS spectra corresponding to peptide ions 870.67, 986.77, 1119.00, 1189.86.

We have developed most of the details of the experimental protocol except for the last step of loading the sample onto SCX and RP column. We expect to complete this remaining step in the next 3-4 weeks.

### **Key research accomplishments**

We have developed an extremely detailed protocol for proteomic analysis of the serum samples. There are no research accomplishments to date (end of year 1).

### **Reportable Outcomes**

None to date

### **Conclusions**

Previous proteomic studies of breast cancer have used lower resolution proteomic approaches in cross-sectional settings only. By using high-resolution proteomic approaches in a prospective setting, this ongoing project may enhance our ability to both identify those women at increased risk of breast cancer and to intervene before they progress to cancer. Furthermore, it may provide insight into the biological processes underlying breast cancer development, for example, by identifying the protein markers underlying case-control proteomic differences and by leading to identification of the associated candidate genes that might be responsible for differences in susceptibility.

### **References**

None

## Appendices

### PROTEMIC PROCEDURES

#### 1. Removal of abundant serum proteins by immunodepletion chromatography

**Column** 6.4 × 63 mm, affinity column from Genway Biotech Inc (San Diego, CA)  
Cat. Number : MIXED12-LC12

**Buffers** Dilution buffer: 10 mM Tris-HCl, 150 mM NaCl, pH 7.4  
Stripping buffer: 0.1 M glycine-HCl, pH 2.5  
Neutralization buffer: 0.1M Tris-HCl, pH 8.0  
Reduction buffer: 500 mM Tris-(2-carboxyethyl)phosphine (TCEP) from Pierce  
Alkylation buffer: 10.5M methyl methane-thiosulfonate (MMTS) from Pierce

#### Instrumentation

AKTApurifier

Loop size (250 µl)

Flow cell (Cat. 18-1147-25) (3mm pathlength, 0.7 µl)

Syringe: 250 µl syringe from Hamilton

#### Procedure

- a. Dilute 20 µl serum samples with 80 µl dilution buffer and spin at 13,000 rpm for 5min.
- b. Fill the loop with dilution buffer, draw up the supernatant with 250 µl syringe and inject the sample into the loop.
- c. Using the depletion method saved on the AKTApurifier \*, start the run.
- d. Collect the flow-through fraction (extinction absorption at 280 nm) with 15 ml capped round-bottomed polyethylene tube, ~3 ml.
- e. Add 1g/ml of GnHCl to the flow-through fraction. Dissolve. The final volume will be about 5.1 ml if 3 ml fraction is used.
- f. To each tube (5.1 ml) add 21µl of 500mM of TCEP, final concentration is 2 mM, vortex to mix, then spin and incubate the tube at 60 °C in water bath for 1 hr, spin to collect condensate.
- g. To each tube, add 3.0 µl of 10.5 M MMTS, vortex to mix, and then spin. Incubate the tubes at room temperature for 30 min.

\* On AKTAPurifier computer, go to method/ depletion method. If computer crashes, use the following parameters.

- a. The column volume (CV) is 2 ml.
- b. Equilibrate the column for 0.1 CV and empty loop for 0.15 CV.
- c. Wash column with dilution buffer at 0.1 ml/min for 1.8 CV then at 0.2 ml/min for 1.5 CV.
- d. Wash column with stripping buffer for 5.0 CV at 1.0 ml/min.
- e. Wash column with neutralization buffer for 2.5 CV at 1.0ml/min then equilibrate column with dilution buffer for 3.0 CV at 1.0ml/min.

## 2. C4 Concentration and Buffer Exchange

**Column:** 2 × 30 mm C<sub>4</sub> reverse phase column from Perkin Elmer  
Catalog number: 0711-0062

### Buffers:

Buffer A: 0.1% HCOOH (pH ≤ 2)

Buffer B: 80% acetonitrile, 10% n-butanol, 0.08% HCOOH

### Instrumentation

AKTApurifier

Loop size (8ml)

Flow cell (Cat. 18-1147-25) (3mm pathlength, 0.7 μl)

Syringe: 10ml syringe (Henke SASS WOLF GMBH), catalog number: 5B14048

SpeedVac

### Procedure

1. Using the method on the AKTA computer (method/ C4 concentration)\*\*, pass the solution through the C<sub>4</sub> reverse phase column. Wash with 0.1 % HCOOH to remove GnHCl, and elute proteins with a steep gradient of 0-90% buffer B, collect eluate at extinction absorption of 280nm, ~ 1.2 ml.
2. Reduce volume of eluate to 10μl-15μl by Speedvac in a 1.5 ml eppendorf tube. Do not dry completely.

\*\* if the computer crashes, using the following parameters

- a. The column volume (CV) is 0.1 ml.
- b. Equilibrate the column for 3 CV and empty loop for 65 CV.
- c. Wash column with buffer A at 0.3 ml/min for 20 CV then at 0.1 ml/min for 2.0 CV.
- d. Wash column with a gradient buffer B (from 0 to 100) for 8.0 CV at 1.0 ml/min and 100% buffer B for 10 CV at 0.1 ml/min.
- e. Wash column with gradient buffer B (from 100 to 0) for 10 CV at 0.5ml/min then equilibrate column with buffer A for 2.0 CV at 0.5ml/min.

### 3. Tryptic Digestion and iTRAQ Labeling

#### Instrument:

Thermomixer R

#### Buffers and Reagents:

Digestion buffer: 0.5 M Triethylammonium biocarbonate (TEAB), from Applied Biosystems

Labeling buffer: 50 mM TEAB in 75% Ethanol

TPCK treated trypsin from Promega, catalog # ( V5111)

#### Procedure:

- a. To each sample, add 90  $\mu$ l digestion buffer, adjust with 1 M KOH if needed, pH should be around 8.5.
- b. For each sample, reconstitute a vial of trypsin (20  $\mu$ g) with 20  $\mu$ l of mill-Q water, vortex to mix, and then spin.
- c. To each sample tube, add 5  $\mu$ l of the trypsin solution, vortex to mix then spin. Discard remaining trypsin.
- d. Incubate the tubes at 37 °C overnight (12-16 hr) in Thermomixer R at 300rpm.
- e. Spin to collect condensation.
- f. Reduce to 10-15  $\mu$ l by SpeedVac and add 100  $\mu$ l of labeling buffer (50 mM TEAB in 75 % ethanol), vortex to mix, and then spin.
- g. Transfer the contents of one iTRAQ reagent vial to one sample tube, vortex each tube to mix, and then spin. Incubate the tubes at room temperature for 1 hour.
- h. Stop the labeling by adding 300  $\mu$ l of distilled water and allowing excess reagent to completely hydrolyze over an additional 30 min at room temperature.
- i. Combine the contents of each iTRAQ Reagent-labeled sample tube into a 2.0 ml eppendorf tube, vortex to mix, and then spin.

### 4. Cation exchange chromatography

#### Column

PolySulfoethyl A <sup>TM</sup> column (2.1  $\times$  100 mm, 5  $\mu$ m, 300Å) from Poly LC Inc

Cat. Number: 102se0503

#### Buffers

Buffer A: 20% acetonitrile, 10 mM KH<sub>2</sub>PO<sub>4</sub>, adjusted to pH 3 with KOH.

Buffer B: 20% acetonitrile, 10 mM KH<sub>2</sub>PO<sub>4</sub>, 700 mM KCl, adjusted pH3 with KOH.

#### Instrumentation

AKTApurifier

Loop (3ml)

Flow cell (Cat. 18-1147-25) (3mm pathlength, 0.7 ul)

Syringe: syringe from Hamilton (2.5ml), catalog number: 7650-01.

Micro Static Mixing Tee Assy: Catalog number (M-540) from upchurch Scientific, not installed.

### **Procedure**

Under development: Load diluted sample onto the column and elute with a gradient of buffer (700 mM KCl, 20% acetonitrile, 10mM KH<sub>2</sub>PO<sub>4</sub>, pH3) over 15 min at a flow rate of 0.1ml/min with fractions collected at 1-min intervals. The gradient need to be optimized.

### **5. Reverse Phase Chromatography-MALDI-TOF/TOF mass spectrometry**

**Column** RP –C<sub>18</sub> capillary column (300 μm ×150 mm) purchased from Dionex Corp

### **Buffers**

Buffer A: 0.1% TFA, 5% acetonitrile, 95% H<sub>2</sub>O

Buffer B: 0.1% TFA, 95% acetonitrile, 5% H<sub>2</sub>O

### **Instrumentation**

Ultimate 3000

ABI 4700 Proteomic Analyzer from Applied Biosystems

### **Procedure**

In development: Performed the experiment with standard peptides, not serum samples as follows:

Transfer samples to well-plate autosampler (WPS) and automatically inject and capture onto a trap column.

Elute onto a RP column from Dionex Corp with a gradient buffer (buffer A, 0.1% TFA, 5% H<sub>2</sub>O; buffer B,

0.1% TFA, 95% acetonitrile, 5% H<sub>2</sub>O). Test what kind of gradient should be used. Column effluent is

mixed automatically in a 2:1 ratio with MALDI matrix (saturated cyano-4-hydroxycinnamic acid) with a

robot MALDI spotting device, spotted directly onto the MALDI plate. MALDI plates were analyzed on an

ABI 4700 Proteomic Analyzer from Applied Biosystems.