BIOMARKER DISCOVERY IN EARLY STAGE BREAST CANCER USING PROTEOMICS TECHNOLOGIES

Guihong Qi

Submitted to the faclty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Biochemistry and Molecular Biology
Indiana University

October 2008

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Mu Wang, Ph.D., Committee Chair

Sonal Sanghani, Ph.D.

Master's Thesis
Committee

Frank A Witzmann, Ph.D.

Jinsam You, Ph.D.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and encouragement of my thesis advisor, Dr. Mu Wang. Under his supervision I chose this topic and began the thesis. My thanks and appreciation go to him for persevering with me as my advisor throughout the time it took me to complete this research and write the thesis. It was a valuable experience working under his guidance. Sincerest appreciation also goes to my committee members, Dr. Sonal Sanghani, Dr. Frank A. Witzmann, and Dr. Jinsam You, for having generously given their time and expertise to improve my work. I thank them for their contribution and their good-natured support.

I would like to thank Monarch LifeSciences for providing facilities and financial support. I also would like to thank Dr. Kerry Bemis for assistance on statistical analysis and also the other members of Monarch LifeSciences. My research experience would not have been successful and enjoyable without support from them.

I cannot end without thanking my family, my husband Xigang Li and my daughter Yingxue, their constant encouragement and love I have relied throughout my time at the Academy. It is to them that I dedicate this work.

ABSTRACT

Guihong Qi

Biomarker Discovery in Early Stage Breast Cancer Using Proteomics Technologies

Among women in the United State, breast cancer is the most common cancer diagnosed in women with approximately 200,000 new cases reported each year and the second leading cause of cancer-related deaths in women, according to the American Cancer Society. Diagnosing breast cancer as early as possible improves the likelihood of successful treatment and can save many lives. However, using mammography as a current method to detect breast tumor has intrinsic limitations. Thus early diagnostic biomarkers are critically important for detection, diagnosis, and monitoring disease progression in breast cancers.

Recently, liquid chromatography (LC) mass spectrometry (MS)-based label-free protein quantification method has become a popular tool for biomarker discovery due to its high-throughput feature and unlimited sample size for quantitative comparison under different biological conditions. In this study, we applied this technology with inclusion of statistical analysis to detect the protein differential expression levels in the plasma samples from the early-stage breast cancer patients. With a combined protein classification and pathway analysis, a panel of potential protein biomarkers has been identified.

The results from this study showed that LC/MS-based label-free protein quantification technology along with bioinformatics analysis provides an excellent

opportunity to help determine biomarker candidates for future validation studies and development of new strategies for early diagnostics and disease treatment.

Mu Wang, Ph.D., Committee Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Breast cancer is the most common type of solid tumor diagnosed in women with approximately 200,000 new cases reported each year in the United States. In 2007, more than 40,000 women died of breast cancer in the United States, making it the second leading cause of cancer-related deaths in women [1]. The chance of developing invasive breast cancer at some point in a woman's life is about 1 in 8. The chance that breast cancer will be responsible for a woman's death is about 1/35 [2]. Breast cancer was one of the first malignancies for which targeted therapy was used to treat a subgroup of the affected population [3]. Diagnosing breast cancer as early as possible improves the likelihood of successful treatment [4], and breast cancer survivors are now the largest group of cancer survivors in the United States [5, 6]. Early detection and prevention of this disease is urgently needed because many patients succumb to advanced diseases as the primary tumor metastasizes to other organs. It is evident that early detection for breast cancer can save many lives [7].

Current methods used to detect breast tumors, either benign or malignant, are primarily based on mammography. However, there are intrinsic limitations to mammography as only 63% of breast cancers are localized at the time of diagnosis [3]. Small lesions are frequently missed and may not be visible, particularly in young women with dense breast tissue [8]. For a breast tumor to be detected in mammography, it must be at least a few millimeters in size. Unfortunately, a tumor of this size already contains several hundred million cells. From the cellular point of view, given the fact that a single cell can lead to the development of a whole tumor, it is already at a late stage when a

tumor is detected by mammography [9]. Third, mammograms have a high rate of false positives, which will result in costly and invasive follow-up tests, including biopsies, of which 75% prove benign [10]. Also, there are distinct subgroups of breast cancer for which specific biological targets have not yet been identified [11]. Biomarkers are critically important tools for detection, diagnosis, treatment, monitoring, and prognosis. Biomarkers are biological molecules that are indicators of physiological state and also of change during a disease process [12]. The value of a biomarker lies in its ability to provide an early indication of the disease and to monitor disease progression.

The primary goal of this study is to discover potential protein biomarker candidates using early stage breast cancer patient samples and provide valuable information for biomarker validation studies, thus developing new strategies for early detection, diagnostics, disease monitoring, and therapeutic treatment. In the previous studies, some potential biomarkers of breast cancer have been suggested [4, 13, and 14]. As these were identified using one-protein-at-a-time approaches, they may or may not be true biomarkers of breast cancer. It is believed that biomarkers are more influential as a panel of proteins within a biological sample—there seems to be a growing consensus that a panel of markers may be able to supply the specificity and sensitivity that individual markers lack [14, 15]. Thus, measurement of multiple proteins in a single assay may give a better and more complete picture of what is happening at the protein expression level that is associated with the disease. In addition, under diseased conditions, it is beneficial to be able to look at multiple proteins to develop a greater understanding of the disease and how it affects life.

Proteomics has become the most powerful and efficient methodology in recent years for simultaneous analysis of thousands of proteins on the basis of differences in their expression levels and post-translational modifications involved in cancer progression [16]. Currently, there is no common consensus within the field as to which proteomic technology can attain complete and quantitative protein coverage of all proteins in a biological sample.

The most commonly used proteomic approach is accomplished by a combination of either two-dimensional gel electrophoresis (2DE) or liquid chromatography (LC) to separate and visualize proteins/peptides and mass spectrometry (MS) to identify, characterize, and quantify them. 2DE has been the workhorse of proteomics for the past decade and is still one of the most widely used tools for separating proteins [17], but its biggest disadvantage is the inability to cover the dynamic range of proteins in a proteome. One alternative strategy to partially overcome the disadvantage of 2DE is LC/MS-based technology, primarily stable isotopic labeling technology coupled with MS. Although some successes using this technology for protein quantification have been reported [18], it is not always practical and has several disadvantages. For example, labeling with stable isotopes is expensive and the isotopic labels sometimes exhibit chromatography shifts that can make quantification of differentially labeled peptides computationally difficult [19]. Moreover there may not be enough different isotopes to allow for simultaneous quantification of proteins from multiple samples (i.e., >8 groups) [19], and it remains technically challenging to characterize the global proteome due to the fact that proteins without cysteine residues cannot be labeled.

More recently, LC/MS-based label-free protein quantification has gradually gained its popularity due to its high-throughput feature and unlimited sample size for quantitative comparison under different biological conditions. It uses extracted ion chromatograms (XICs) from mass spectrometric analysis for relative quantitation of protein expression [16, 20, 21, and 22].

The focus of this project is to use the label-free protein quantification platform to compare plasma proteins from early stage (stage I and stage II) breast cancer patients in order to identify biomarkers for early detection of breast cancer. Using a large sample set (80-sample) will not only allow us to identify potential breast cancer biomarker candidates, but also establish an optimized platform and protocol for biomarker discoveries. Information obtained from this study will also help to determine biomarker candidates for future validation studies and development of new strategies for diagnostics and disease treatment.

2. Materials and Methods

2.1 Human Plasma Samples

Forty plasma samples from women with breast cancer and 40 plasma samples from healthy age-matched volunteer women (control) were collected by the Hoosier Oncology Group (HOG) (Indianapolis, IN, USA). All patients involved in this study were diagnosed with a stage II or earlier breast cancer. Details of these patients are shown in Table 1.

2.2. Experimental Designs

The study is consisted of two groups of plasma samples, 40 plasma samples from women with stage I or II breast cancer (all prior to chemotherapy) and 40 plasma samples from healthy age-matched women to serve as controls. Single injections for each sample were performed. The tables shown below summarize the patient information and the experimental design.

Table 1: Summary for 80 samples based on age ranges

Age Distribution	30-39		40-65		> 65	
	Cancer	Healthy	Cancer	Healthy	Cancer	Healthy
Number of patients	3	3	29	34	8	3
Race	3-White	1	28-White	31	8-White	3
	0-Black	1	1-Black	3	0-Black	
	0-Others	1	0-Others		0-Others	
Ethnicity	3-Nonhispanic	3	28-Nonhispanic	33	8-Nonhispanic	3
	0-Hispanic		1-Hispanic		0-Hispanic	
	0-Unknown		0-Unknown	1	0-Unknown	
Metastasis	3-No		24-No		6-No	
	0-Yes		1-Yes		1-Yes	
	0-Unknown		4-Unknown		1-Unknown	
Cancer Type*	2-INV		12-INV		4-INV	
	1-DCIS		10-DCIS		4-DCIS	
	0-Unknown		7-Unknown		0-Unknown	
Tumor Size (cm)	m = 1.3 [0.2,1.7]		m = 2.37 [0,5.5]		m = 1.06 [0,4.5]	
Lymph Node+	0		10		4	

*INV: invasive; DCIS: ductal carcinoma in situ

Table 2: Experimental Design

Group	Condition	Number of Samples	Number of Injections
0H	Healthy	40	40
1C	Cancer	40	40

2.3 Materials

Ammonium carbonate, ammonium bicarbonate, urea, formic acid, lysozyme, 2-Iodoethanol, and triethylphosphine were all purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile and MS grade water were purchased from Honey Burdick & Jackson (Morristown, NJ, USA). Trypsin was purchased from Worthington Biochemical Corporation (Lakewood, NJ, USA). Seppro tip IgY-12 and reagent kit were purchased

from GenWay Biotech (San Diego, CA, USA) and HPLC column – Xbridge-C18 (2.1 mm x 50 mm, pore size = 2.5 µm) was purchased from Waters (Ireland).

2.4 High Abundant Protein Removal

A large number of proteins present in blood plasma indicate an excellent biospicemen for discovering biomarkers for potential clinical diagnostics and therapeutics. However, low-abundance proteins are often undetectable in proteomic analysis of plasma due to the high abundance of some circulating proteins [23]. These high-abundance plasma proteins are the main cause of assay background. For example, albumin, the most abundant protein in plasma, constitutes over half of the plasma proteins and is present at 30-50 mg/mL concentration. In contrast, most of the potential biomarkers are secreted into the blood stream at very low copy number, especially in the early onset of diseases [24, 25]. Thus, removal of the high-abundance proteins is a critical step in biomarker discovery. In this study, we used the GenWay Seppro Tip IgY-12 and PSS Bio Instrument's automated Magtration System 12 GC to remove the top 12 most abundant proteins in plasma. Our data showed that the GenWay Seppro Tip IgY-12 system has both efficiency and reproducibility required for biomarker discovery when compared with several other commercially available abundant protein removal kits, including Montage (Millipore) and Multiple Affinity Removal System (Agilent). The performance of this Tip IgY-12 also has been reported to be specific, efficient, and reproducible in a previous study [23]. The Seppro Tip IgY-12 is packed with immobilized IgY antibody beads for immunoaffinity capture of human albumin, IgG, α1-antitrypsin, IgA, IgM, Transferrin, Haptoglobin, α1-acid glycoprotein, α2-Macroglobulin, HDL (Apolipoprotein A1 & AII), and Fibrinogen [26]. After the high abundant proteins

removal, the low abundant proteins in the flow-through fractions were analyzed. The Seppro Tip products are designed to be used with PSS Bio Instrument's automated Magtration System 12GC. Twelve tips are simultaneously operated to process twelve samples at once. Specific removal of 12 high-abundance proteins depletes approximately 95% of total protein mass from human plasma [26].

For this study, 80 human plasma samples were centrifuged at 10,000 rpm for 1 minute to remove insoluble material, and the clear supernatant was used for downstream processing. Briefly, 15 µL clear human plasma samples were diluted with TBS buffer (10 mM Tris-HCl, 0.15 M NaCl, pH 7.4) to a final volume of 500 µL in a 1.5 mL screw-cap tube. The sample containing tubes, eluting buffer tubes (0.25 M Glycine-HCl, pH 2.5), washing buffer tubes (TBS, 10 mM Tris-HCl, 0.15 M NaCl, pH 7.4), neutralization buffer tubes (0.25 M Tris-HCl, pH 8.0) and depletion tips were all loaded on the PSS Bio Instrument's automated Magtration System 12GC before the depletion protocol started. The flow-through (depleted) fractions were collected, and the bound fractions containing high abundant proteins can be recovered with elution buffer if desired. The column was then washed with washing buffer and re-equilibrated with neutralization buffer for application of subsequent samples. This column can be reused for 25 cycles.

2.5 Protein Reduction, Alkylation and Digestion

The protein concentration of the collected flow-through fractions were determined by the Bradford protein assay [33]. The collected flow-through fractions were then concentrated to about 30 µL from 500 µL with a spin concentrator (Barnstead/Genevac, Genevac LTD, IPSwich England) and spiked with 0.15 µg chicken lysozyme (which was used as internal standard for QA/QC purpose). 30 µL of 8 M urea, 25 µL of water, and 5

µL of 1 M ammonium carbonate, pH 11.0, were then added to the depleted plasma samples. Next, an equal volume (80 µL) of reduction/alkylation cocktail (2% iodoethanol, and 0.5% triethylphosphine in acetonitrile) was added [27]. The solutions were capped and incubated for 1.5 hrs at 37°C, after which it was dried overnight using a speed-vacuum. The pellet was then dissolved in 150 µL of a trypsin solution (0.6 µg trypsin in 100 mM ammonium bicarbonate, pH 8.0) to produce a final concentration of 1.6 M urea solution. The digestion was carried out at 37°C overnight. 100 µL (20 µg) of this digest was then injected onto a Surveyor HPLC system coupled with an LTQ mass spectrometer [Thermo-Fisher Scientific, Waltham, MA, USA] in a random order.

2.6 Mass Spectrometry Instrumentation

All tryptic digests were separated by an XBridge (2 mm x 50 mm) C-18 reversed phase column (Waters, Milford, MA, USA) at a flow rate of 200 µL/min. The linear gradient conditions for elution of peptide were 10-95% of 0.1% formic acid in 50% acetonitrile (Buffer B) over 120 min, followed by 5 min at 100% of 0.1% formic acid in 80% acetonitrile (Buffer C), then followed by 90% of 0.1% formic acid in water (Buffer A) and held for 17 min. Between each sample in the set an injection of water is made and a shortened (60 min) gradient is performed to reduce carryover. The effluent from HPLC column was directly electro-sprayed into the LTQ mass spectrometer. The LTQ was performed in positive ion mode with 4.8 kV electrospray potential, a sheath gas flow of 20 arbitrary units, and a capillary temperature of 225°C. The source lenses were set by maximizing the ion current for the $M+2H^+$ charge state of angiotensin and data were collected in triple-play mode (MS scan, Zoom scan, and MS/MS scan) with *m/z* range of 350-2000 amu.

2.7 Peptide and Protein Identification

All data collected from triple-play experiment were used to estimate the quality of subsequent monoisotopic and average mass of the peptide, the charge state, and MS/MS spectra of the peptide (shown in Figure 2.7.1). Protein identification was carried out using the software package licensed from Eli Lilly and Company [20]. To minimize false-positive identifications, the low quality data were filtered out by the same software package [20]. Briefly, filtered data were subsequently searched against the IPI (International Protein Index) and the Non-Redundant (NCBI) databases using both the SEQUEST and X!Tandem algorithms. Proteins identified by SEQUEST and X!Tandem are categorized into priority groups based on the quality of the protein identification as shown in Table 3. The Peptide ID confidence assigns a protein to a 'HIGH' or 'MODERATE' classification based on the peptide with the highest peptide ID Confidence (the best peptide). Proteins whose best peptide has a confidence between 90-100% are assigned to the 'HIGH' category regardless of whether there are other peptides having low confidence. Proteins whose best peptide has a confidence between 75-89% are assigned to the 'MODERATE' category. All peptides with confidence less than 75% are filtered out by the software before further analysis. To confirm protein identification, each database search result was then searched against a reverse database. If any MS/MS spectra were matched against the reverse database, it was then excluded from the list.

Table 3: Classification of protein identification

Protein Priority	Peptide ID Confidence	Multiple Unique Sequences
1	High (90-99%)	Yes
2	High (90-99%)	No
3	Moderate (75-89%)	Yes
4	Moderate (75-89%)	No

The confidence in protein identification increases with the number of distinct amino acid sequences identified. Therefore proteins are also categorized depending on whether only one or multiple unique peptide sequences are obtained. A protein is classified as 'YES' in the 'Multiple Sequences' column if it has at least two distinct amino acid sequences with the required ID confidence; otherwise it is classified as 'NO'. Priority assignments reflect the level of confidence in the protein identification. Priority 1 proteins would have the highest likelihood of correct identification and Priority 4 proteins the lowest. This priority system is based on the quality of the amino acid sequence identification (Peptide ID Confidence) and whether one or more sequences are identified (Multiple Sequences). We typically view any protein identification outside of priority 1 as questionable [29]. All data processing is carried out on a Linux cluster using highly parallel processing and data qualification and filtering software.

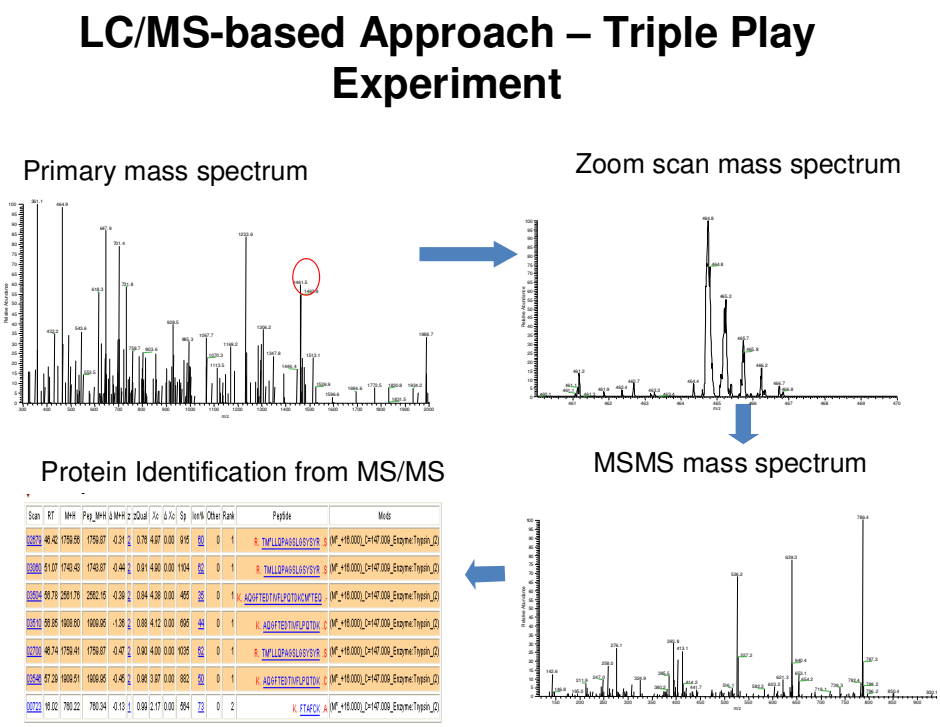

Figure 2.7.1 The triple-play experiment for label-free protein quantification

2.8 Peptide and Protein Quantification

Protein quantification was also carried out using the same software package we licensed from Lilly as described earlier [20]. Briefly, once the raw files are acquired from the LTQ, all extracted ion chromatograms (XIC) are aligned by retention time (Figure. 2.8.1). To be used in the protein quantification procedure, each aligned peak must match parent ion, charge state, daughter ions (MS/MS data) and retention time. After alignment, the area-under-the-curves (AUC) for individually aligned peaks from identified peptides from each sample are computed; the AUCs are then compared for relative protein abundance.

One of the key features of the algorithm for protein quantification is the chromatographic peak alignment, because large biomarker studies can produce chromatographic shifts due to multiple injections of the samples onto the same HPLC column. Un-aligned peak comparison will result in larger variability and inaccuracy in peptide quantification [20]. A graphical example of a comparison of peptide quantities across a complex biological sample is shown in Figure. 2.8.1. All peak intensities are transformed to a $\log_2$ scale before quantile normalization [28]. Quantile normalization is a method of normalization that essentially ensures that every sample has a peptide intensity histogram of the same scale, location, and shape. This normalization procedure removes trends introduced by sample handling, sample preparation, total protein differences, and changes in instrument sensitivity while running multiple samples. If multiple peptides have the same protein identification, then their quantile normalized $\log_2$ intensities are averaged to obtain $\log_2$ protein intensities. The $\log_2$ protein intensity is the final quantity that is statistically modeled. A separate model is fit for each protein. The appropriate model depends on the phenotype associated with the protein expression. Phenotypes with categorical response would probably be studied with an ANOVA model whereas phenotypes with numerical response would be studied with a regression model. Significance is first measured by a *p*-value. All p-values are then adjusted to control for the False Discovery Rate (FDR). The FDR is estimated by the *q*-value which is an adjusted *p*-value. The FDR is the proportion of significant changes that are false positives. If proteins with a $q\text{-value} \leq 0.05$ are declared significant, it is expected that 5% of the declared changes will be false positives. A data processing flow chart is shown in Figure. 2.8.2.

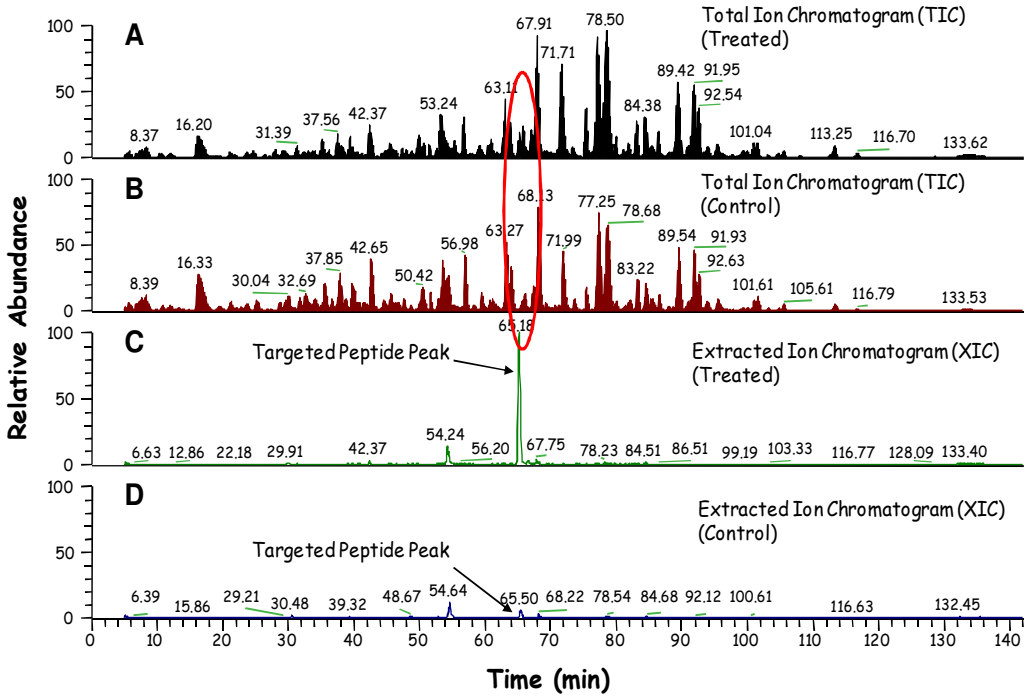

Figure 2.8.1 Peptide quantification by extracted ion chromatograms (XICs). Panels A and B are total ion chromatograms (TICs) from treated and control sample respectively. Panels C and D are extracted ion chromatograms (XICs) from treated and control samples. The area-under-the-curve (AUC) can be calculated and compared for the relative quantity of the peptide of interest (indicated by arrows), thus protein of interest.

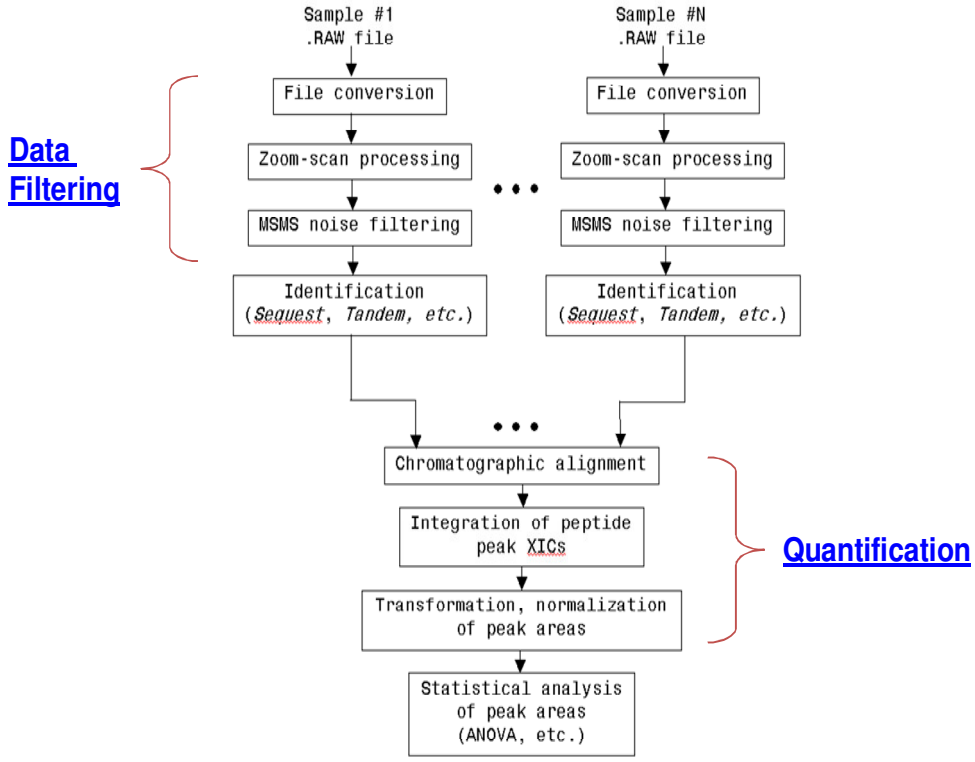

Figure 2.8.2 Data processing flow chart

2.9 Quality Assurance and Quality Control

In this experiment, all of the samples were prepared by the same person. All injections were randomized and performed using the same C18 microbore column. All buffers were prepared at the same time for all injections. To assess the stability of the column and instrument, the same amount of chicken lysozyme was spiked into every sample before tryptic digestion. The spiked internal standard chicken lysozyme can help check ion intensities before and after normalization and served as a QA/QC standard. In the plot shown in Figure 2.9.1, the individual protein quantities (peak intensities) are displayed for each injection. The overall mean for each group is displayed by the line

across the plot. Since a constant amount of chicken lysozyme was spiked into the entire sample, it should show no significant changes between groups. If there is a significant group change then it is advisable to be cautious when interpreting significant changes in other proteins with smaller fold changes.

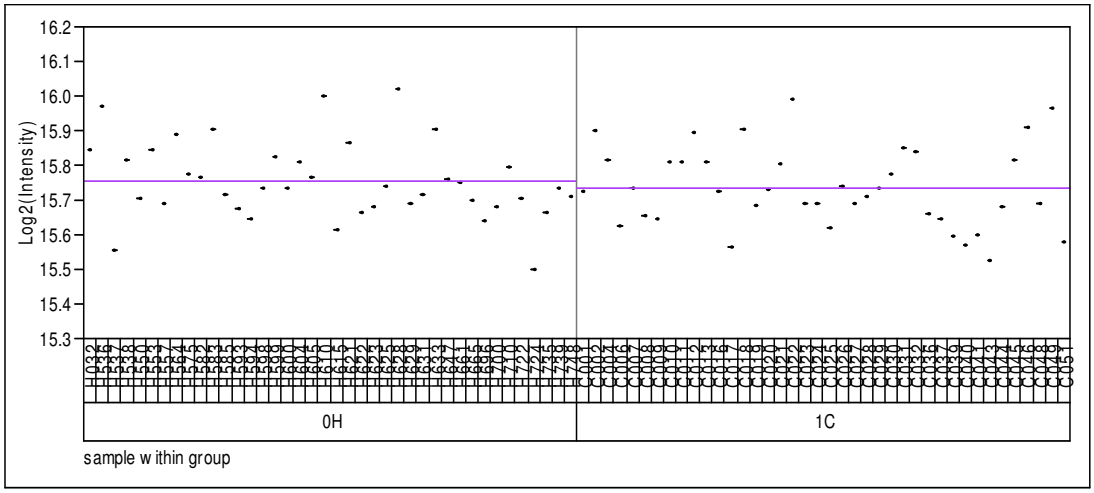

Figure 2.9.1 The individual protein intensities for chicken lysozyme are plotted on a $\log_2$ scale. The overall mean for each group is displayed by the line across the plot.

2.10. Pathway Analysis

After the proteins with significant changes between breast cancer and normal control were identified by the LC/MS-based quantitative analysis, the pathway analysis was performed using Pathway Studio™ software (5.0, Ariadne Genomics, Rockville, MD, USA). The differentially expressed proteins were run against the ResNet database that was equipped with functional relationships from other scientific literature and commercial databases. The filters we used included "all shortest paths between selected entities" and "cell process". Protein interactions and their biological processes were reviewed. A list of proteins of interest was generated from this information, including their pathways and functions.

3. Results

3.1 Protein Identification

In this study, with analysis of 40 plasma samples from breast cancer patients and 40 plasma samples from healthy controls, a total of 1422 proteins and 6457 peptides were identified and quantified (summarized in Table 4).

Of these, 501 proteins were identified with high confidence (priority 1 and 2), and 385 proteins showed a significant expression change between cancer patient and healthy control (false discovery rate less than 5%). The median %CV for priority 1 protein was 14.24% (technical plus biological variations), and the overall Median %CV for all proteins was 19.42%. Among the 921 proteins that were less confidently identified (Priorities 3 and 4), there were also 251 proteins that had significant changes.

Table 4: Summary information of the study using LC/MS-based label free protein quantification method

Protein Priority	Peptide ID Confidence	Multiple Sequences	Number of Proteins	Number Significant Changes	Maximum Absolute Fold-change	Median %CV sample
1	High	Yes	222	58	1.4	14.2
2	High	No	279	76	1.4	20.1
3	Moderate	Yes	53	13	1.2	14.3
4	Moderate	No	868	238	1.5	20.7
Overall			1422	385	1.5	19.4

3.2 Protein Quantification

For protein quantification, every peptide quantified had an intensity measurement for each sample. The intensity measurement is a relative quantity giving the AUC from the extracted ion chromatogram (XIC) after background noise removal. The AUC was

measured at the same retention time for each sample after the sample chromatograms had been aligned [20]. The example alignment result of this study is shown in Figure 3.2.1. The intensities were then transformed to the log scale and quantile normalized [28]. If multiple peptides had the same protein identification then their quantile normalized log base 2 intensities were averaged to obtain log base 2 protein intensities. The log base 2 protein intensity is the final quantity that is fit by a separate Analysis of Variance (ANOVA) statistical model for each protein. Figure 3.2.2 shows an example of relative protein expression levels when comparing cancer sample group with control sample group.

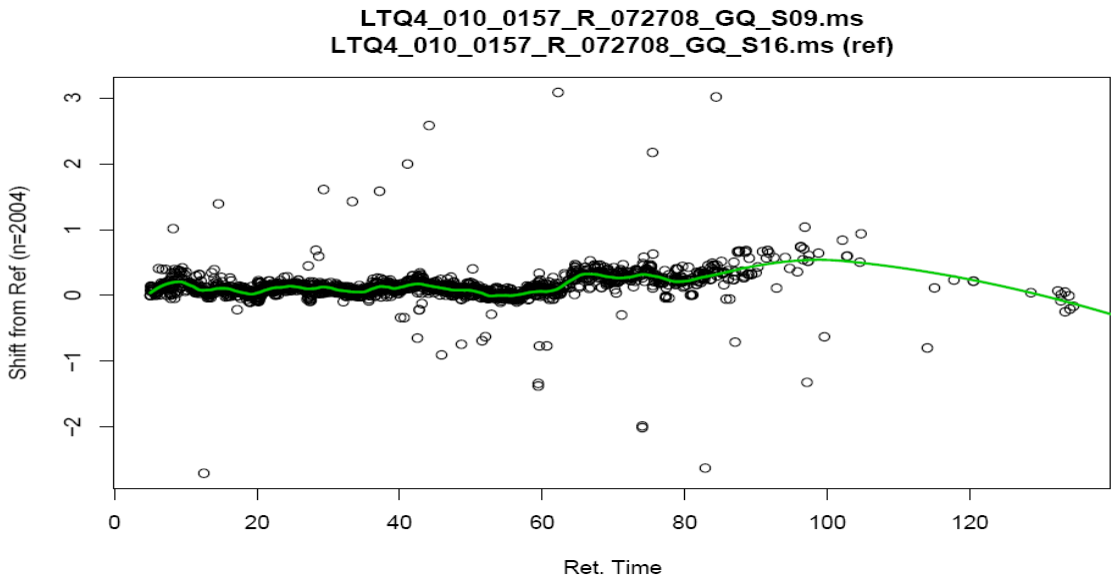

Figure 3.2.1 The extracted ion chromatograms (XIC) is aligned among all samples in the study and a selected reference sample in the study by retention time. To be used in the protein quantification procedure, each aligned peak between the two samples must match parent ion, daughter ion, and charge state and the retention time. A time shifting function puts the samples on the same time scale (in 1 min).

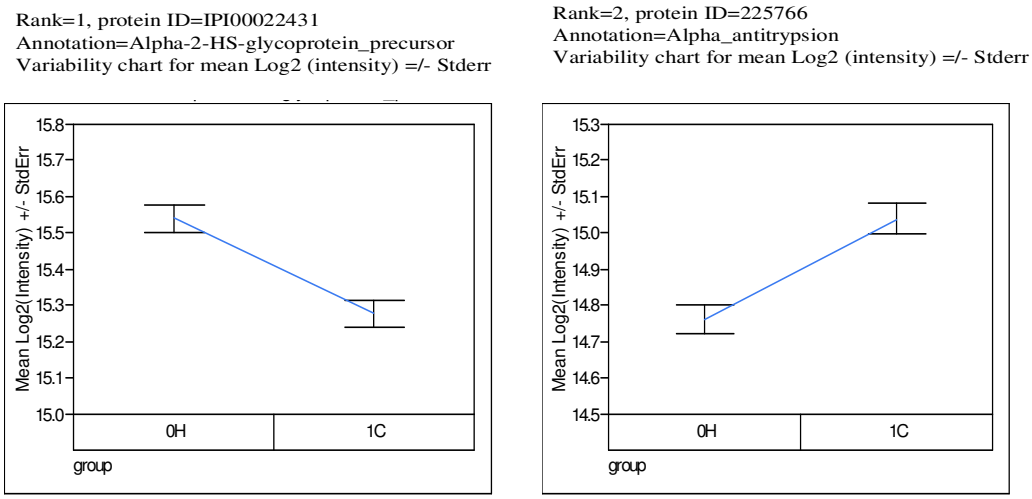

Figure 3.2.2 Example of relative protein expression levels under different conditions. The intensities which are given by the AUC from the XIC are transformed to the log base 2 scale; base 2 is popular because a two-fold change is transformed to a unit change on a log base 2 scale. Error bars show standard errors based on the ANOVA model. Rank is assigned by sorting all the proteins in the order of significant change (Yes, No), priority (1-4), and q value.

3.3 **Analysis**

A significant fold change between groups is based on controlling the false discovery rate (FDR) at less than 5%. The FDR is estimated by the q-value which is an adjusted p-value. The FDR is the proportion of significant changes that are false positives. If proteins with a q-value less than 5% are declared significant, that means the chance of false positives are less than 5%. Because protein intensity is on a log base 2

scales, the group means and their differences are converted to arithmetic means and fold change as calculated below:

T = Cancer group average of log base 2 scale protein intensities

C = Health control group average of base 2 protein intensities

Fold change = Mean_T / Mean _C when Mean_T ≥ Mean_C (up-regulation)

Fold change = - Mean_C / Mean_T when Mean_C > Mean_T (down-regulation)

Fold change = 1 shows no change

3.4 Genome Ontology Classification of the Detected Proteins with Significant Changes

All proteins (priority 1-4) with significant changes ($q < 0.05$) were annotated and categorized based on their biological function, molecular function, and cellular component with Gene Ontology [31, 32]. Figures 3.4.1, 3.4.2, and 3.4.3 present all proteins with significant changes that were presented in the form of a pie chart. In order to keep the graph less cluttered, only a few of the top ranking ones are included in the pie chart. The next three graphs in Figures 3.4.4, 3.4.5, and 3.4.6, respectively, are for fold-change comparison between groups 0H and 1C. The data we obtained by LC/MS label-free quantification method is graphed to show a better picture of up- and down-regulations of the different proteins with respect to their biological process, molecular functions, and cellular locations. Positive columns represent the number of proteins which are up-regulated in the first group (0H) as compared to the second group (1C) (fold-change value is positive). Negative columns represent the number of proteins which are down-regulated in the first group as compared to the second group (fold-change value is negative)

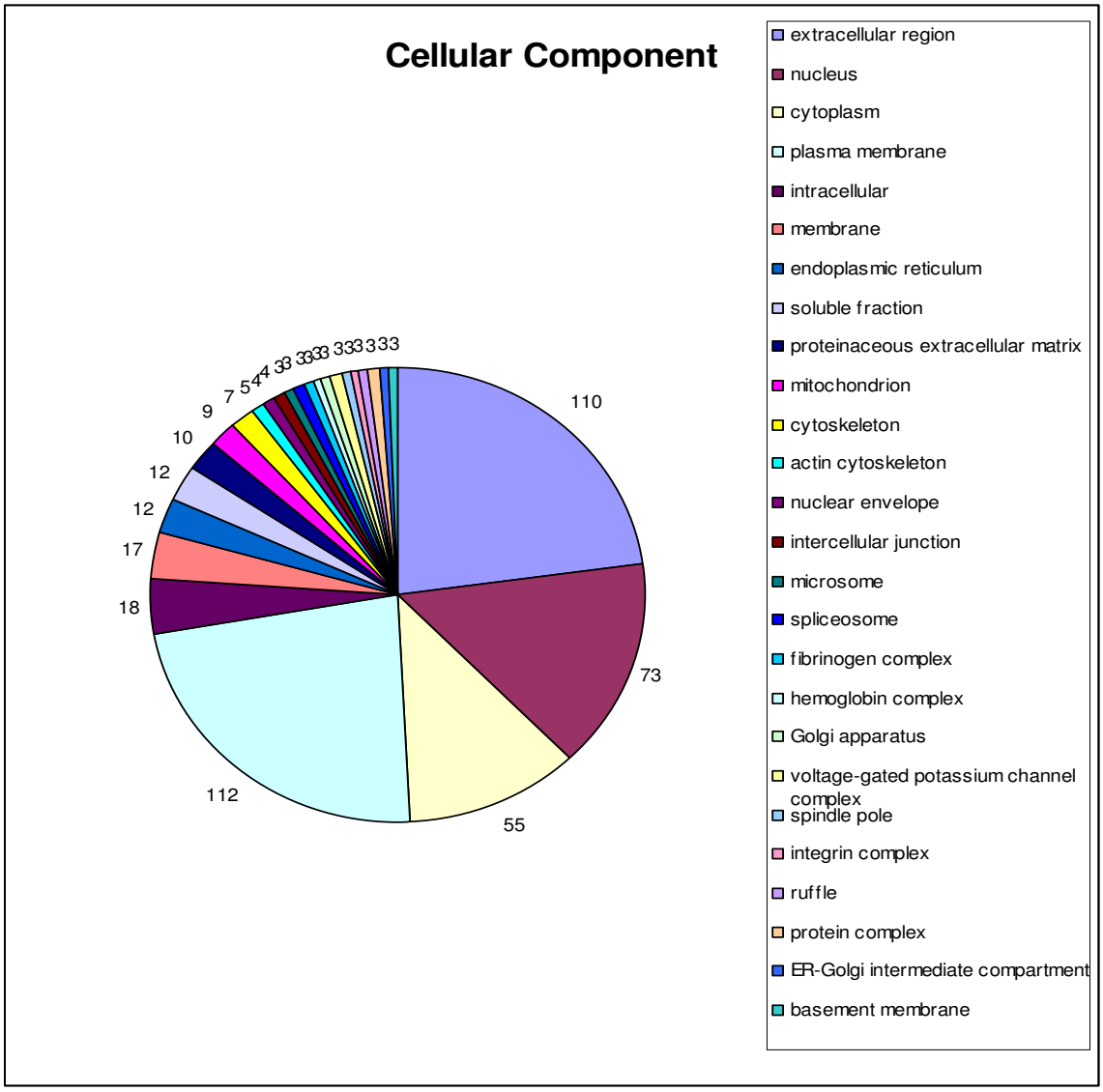

Figure 3.4.1 Cellular Component GO term

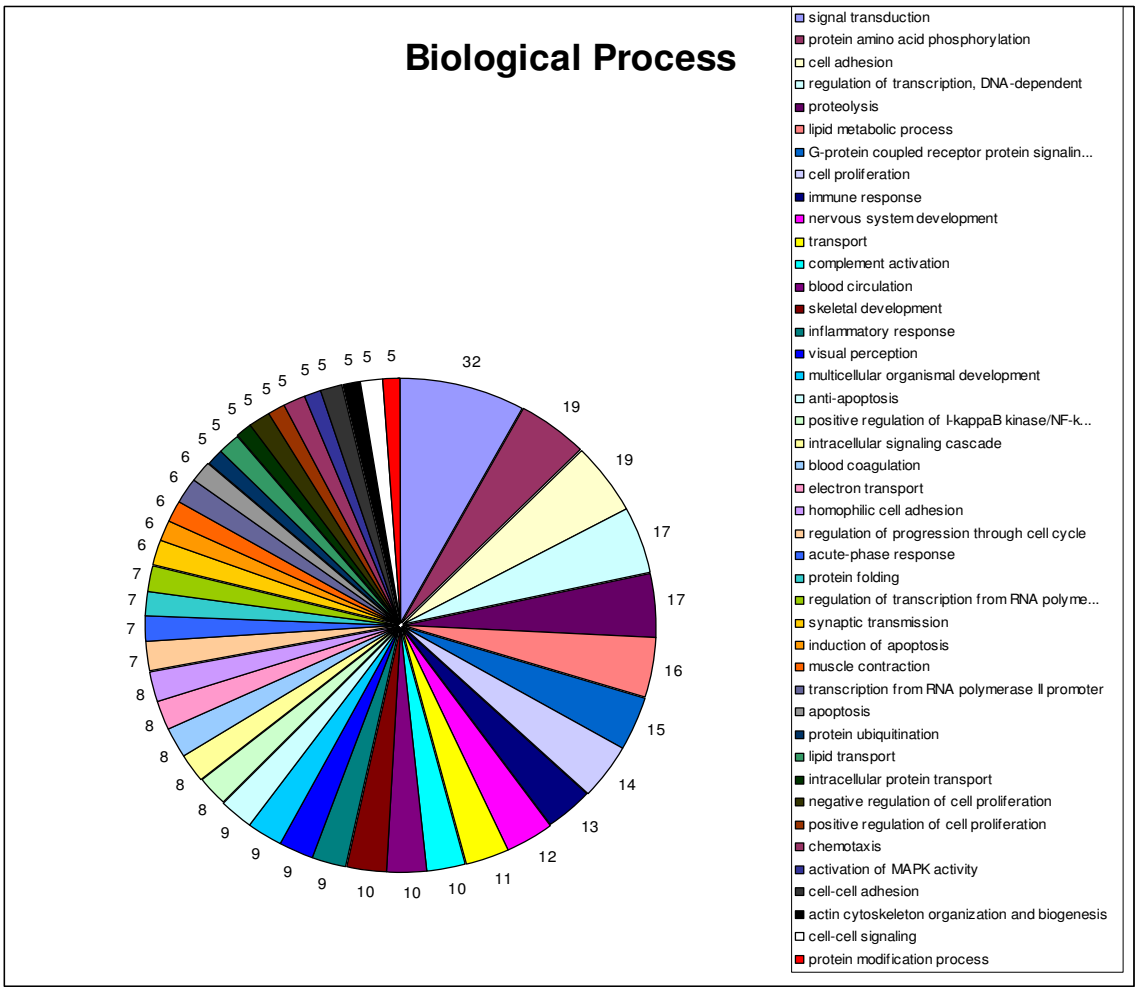

Figure 3.4.2 Biological Process GO terms

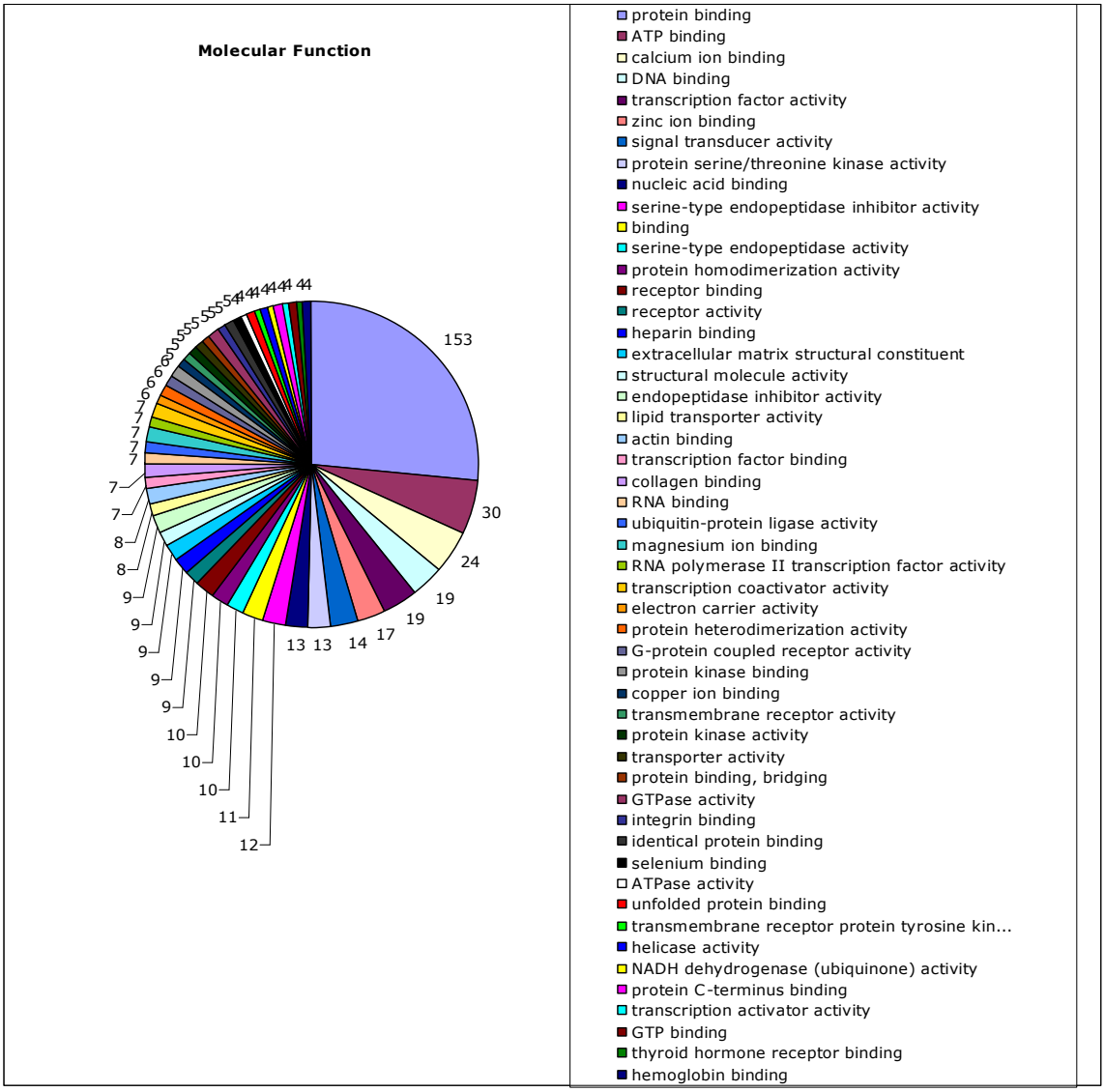

Figure 3.4.3 Molecular Function GO term

- The above three pie charts are for the protein classification with Gene Ontology (GO).
- All proteins with significant changes were categorized based on their biological function, molecular function, and cellular component with GO.
- In order to keep the graph less cluttered, only a few of the top ranking proteins are included in the pie chart.

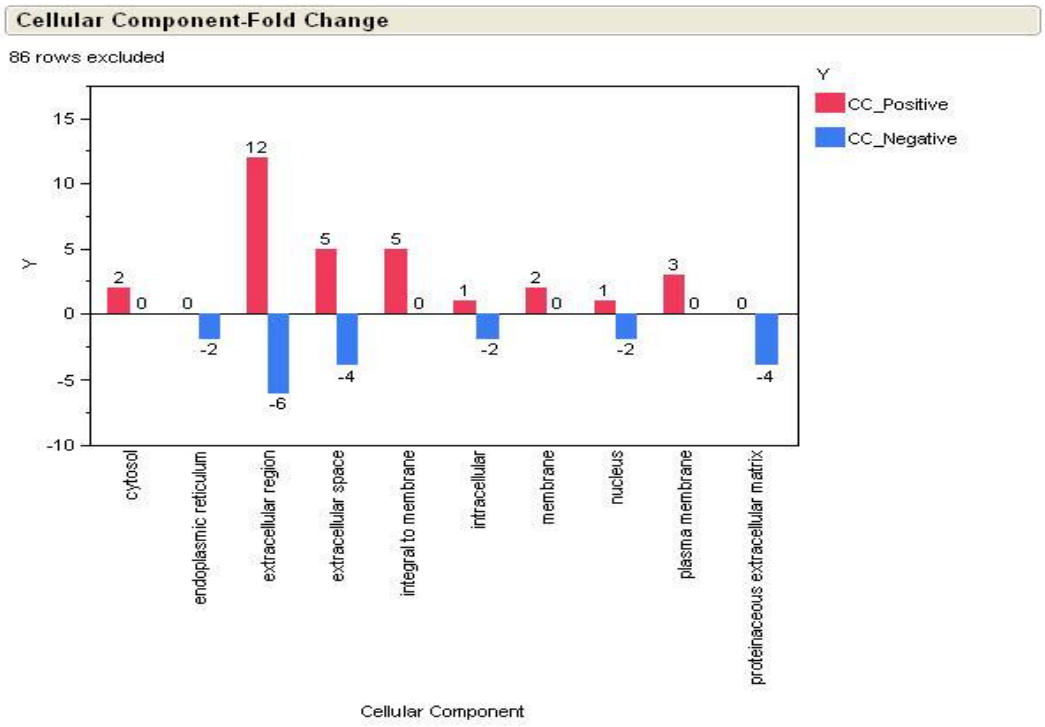

Figure 3.4.4 Classification based on GO Term: Cellular Component

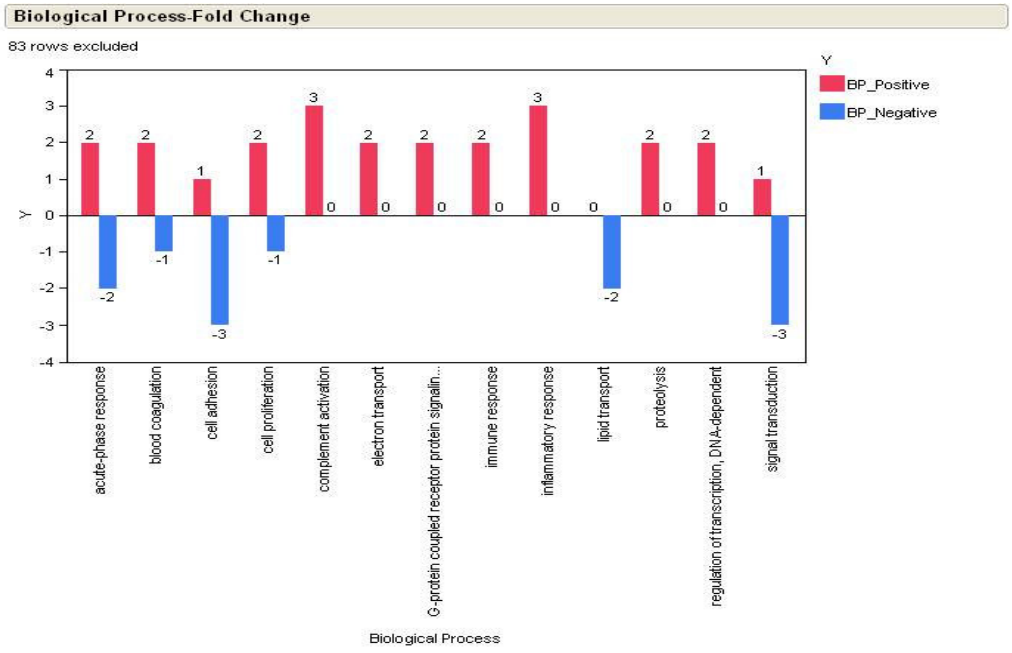

Figure 3.4.5 Classification based on GO Term: Biological Process

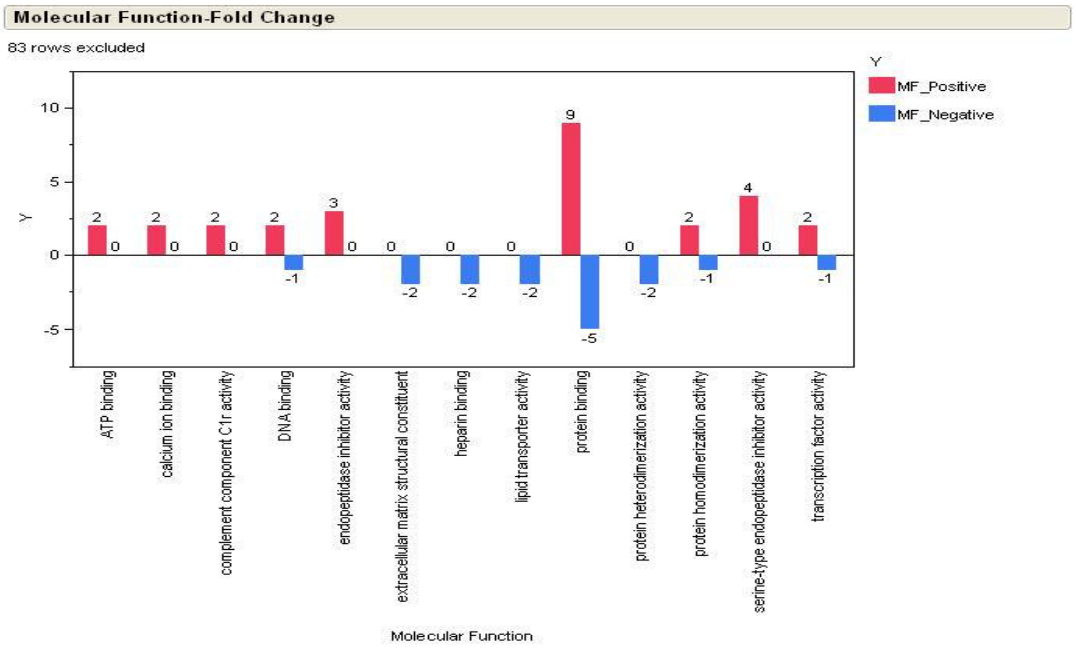

Figure 3.4.6 Classification based on GO Term: Molecular Function

- The above three graphs are for Fold Change comparison between groups 0H and 1C.
- All proteins with significant change were selected.
- Positive column represents the number of proteins which are up regulated in the first group (Healthy) compared with second group (Cancer) (fold change value is positive).
- Negative column represents the number of proteins which are down regulated in the first group compared with second group (fold change value is negatively).

3.5 Comparison with a List of Candidate Cancer Biomarkers

We compared the proteins with significant changes from our data against a list of previously published 1261 candidate cancer biomarkers [14], of which 22 proteins were overlapped (shown in Table 5). A list of 1261 proteins believed to be differentially expressed in human cancer has been compiled from literature and other sources. These

proteins, only some of which have been detected in human plasma, represent a population of candidate plasma biomarkers that could be useful in early cancer detection and monitoring given sufficiently sensitive and specific assays. Most of them have been detected in studies of tissue or nuclear components (tissue, DNA, or RNA). Among these candidates, only few have been validated and approved [14]. This list of cancer biomarkers are only the candidates which were provided for future validation.

Table 5: 22 proteins with significant changes which also present in the published list of cancer biomarker

Gene name	Annotation	Function
AHSG	Alpha-2-HS-glycoprotein_precursor	Function: Promotes endocytosis, possesses opsonic properties and influences the mineral phase of bone. Shows affinity for calcium and barium ions.
APOA2	Apolipoprotein	Function: May stabilize HDL (high density lipoprotein) structure by its association with lipids, and affect the HDL metabolism.
SERPINA3	Isoform_2_of_Alpha-1-antichymotrypsin_precursor	Function: Although its physiological function is unclear, it can inhibit neutrophil cathepsin G and mast cell chymase, both of which can convert angiotensin-1 to the active angiotensin-2.
	Alpha-1-antitrypsin_precursor	Function: Inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin and thrombin. The aberrant form inhibits insulin-induced NO synthesis in platelets, decreases coagulation time and has proteolytic activity agaisnt insulin and plasmin.
FINC	Isoform_1_of_Fibronectin_precursor	Function: Fibronectins bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. Fibronectins are involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape. Interaction with TNR mediates inhibition of cell adhesion and neurite outgrowth (By similarity).
PLG	Plasminogen_precursor	Function: Angiostatin is an angiogenesis inhibitor that blocks neovascularization and growth of experimental primary and metastatic tumors in vivo.
C7	complement_C7_[Homo_sapiens]	Function: C7 is a constituent of the membrane attack complex. C7 binds to C5b forming the C5b-7 complex, where it serves as a membrane anchor.

ORM1	Alpha-1-acid_glycoprotein_1_precursor	Function: Appears to function in modulating the activity of the immune system during the acute-phase reaction.
C6	complement_component_6_[Homo_sapiens]	Function: Involved in the formation of the lytic C5b-9m complex.
IGF2	Isoform_1_of_Insulin-like_growth_factor_II_precursor	Function: The insulin-like growth factors possess growth-promoting activity. In vitro, they are potent mitogens for cultured cells. IGF-II is influenced by placental lactogen and may play a role in fetal development.
SERPINF1	Pigment_epithelium-derived_factor_precursor	Function: Neurotrophic protein; induces extensive neuronal differentiation in retinoblastoma cells. Potent inhibitor of angiogenesis. As it does not undergo the S (stressed) to R (relaxed) conformational transition characteristic of active serpins, it exhibits no serine protease inhibitory activity.
S100A9	Protein_S100-A9	Function: Expressed by macrophages in acutely inflammated tissues and in chronic inflammations. Seem to be an inhibitor of protein kinases. Also expressed in epithelial cells constitutively or induced during dermatoses. May interact with components of the intermediate filaments in monocytes and epithelial cells.
SAA1	SAA1_protein_[Homo_sapiens]	Function: Major acute phase reactant. Apolipoprotein of the HDL complex.
VIL1	Villin-1	Function: Ca (2+)-regulated actin-binding protein.
XRCC1	DNA-repair_protein_XRCC1	Function: Corrects defective DNA strand-break repair and sister chromatid exchange following treatment with ionizing radiation and alkylating agents.
TXLNA	hypothetical_protein_[Homo_sapiens]	Function: May be involved in intracellular vesicle traffic and potentially in calcium-dependent exocytosis in neuroendocrine cells.
BIRC6	Baculoviral_IAP_repeat-containing_protein_6	Function: May protect cells from undergoing apoptosis.
COL11A1	Isoform_B_of_Collagen_alpha-1(XI)_chain_precursor	Function: May play an important role in fibrillogenesis by controlling lateral growth of collagen II fibrils.
NAIP	Baculoviral_IAP_repeat-containing_protein_1	Function: Prevents motor-neuron apoptosis induced by a variety of signals.
ANXA11	Annexin_A11	No annotated function present
ITGA5	Integrin_alpha-5_precursor	Function: Integrin alpha-5/beta-1 is a receptor for fibronectin and fibrinogen. It recognizes the sequence R-G-D in its ligands. In case of HIV-1 infection, the interaction with extracellular viral Tat protein seems to enhance angiogenesis in Kaposi's sarcoma lesions.

FADD	FADD_protein	Function: Apoptotic adaptor molecule that recruits caspase-8 or caspase-10 to the activated Fas (CD95) or TNFR-1 receptors. The resulting aggregate called the death-inducing signaling complex (DISC) performs caspase-8 proteolytic activation. Active caspase-8 initiates the subsequent cascade of caspases mediating apoptosis.

3.6 Pathway Analysis

385 proteins with significant changes from LC/MS data were analyzed using Pathway Studio™ 5.0. A corresponding gene list was created from these proteins. This software was developed to navigate and analyze biological pathways, gene regulation networks and find relationships among genes, proteins, cell processes, and diseases from a dataset. Several proteins were selected based on our data from LC/MS and information obtained from the pathway analysis and other literature searches, which may serve as a panel of biomarker candidates in early stage of breast cancer.

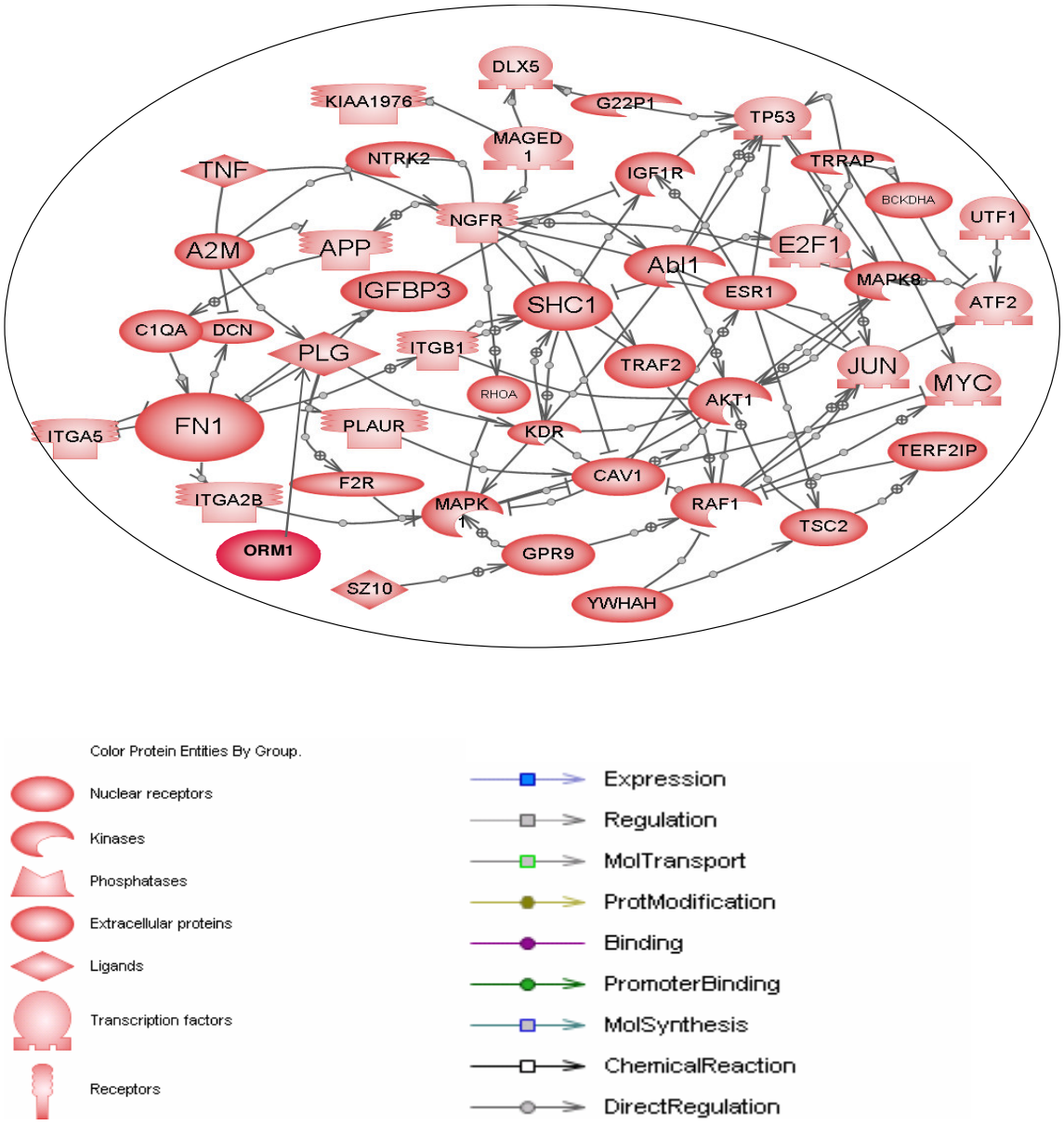

Figure 3.6.1 Pathway Analysis 1: A suggested protein network involving early stage breast cancer, the gene list was run against the ResNet database. The filters were set up including "all shortest paths between selected entities" and "proteins with direct regulation." A few lines were selected for estimating the breast cancer biomarker candidates.

Line 1: ITGA5 → FN1 → IGFBP3 → IGF → TP53 → Breast Cancer

Line 2: SHC1 → IGF → TP53 → Breast Cancer

Line 3: TP53 → ESR1 → TSC2 → Breast Cancer

Line 4: ORM1→PLG → IGFBP3 → IGF → TP53 → Breast Cancer

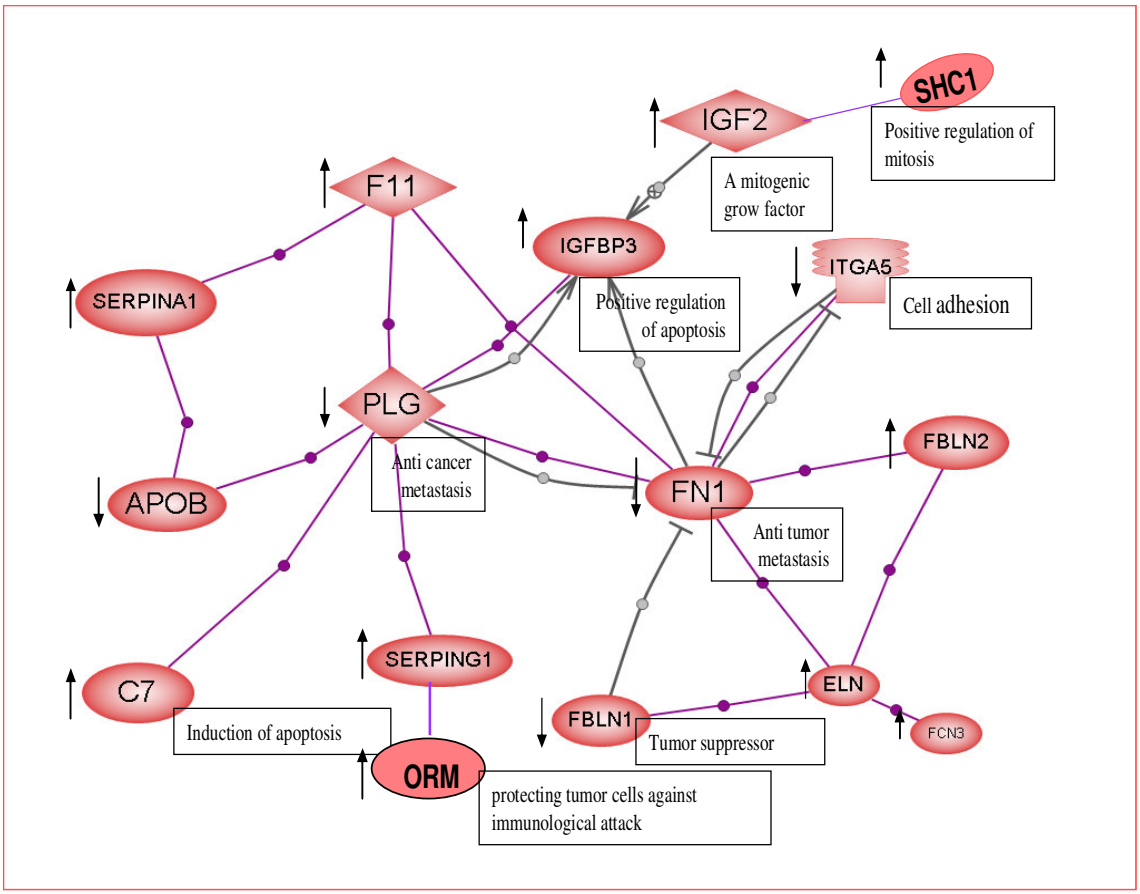

Figure 3.6.2 Pathway Analysis 2: A suggested protein network involving early stage breast cancer, the gene list was run against the ResNet database. The filters were set up including "all shortest paths between selected entities" and "cell process." The functions of the main genes are marked.

Table 6 below shows the proposed biomarker candidates in the early stage breast cancer found by us and they are supported by the pathway analysis and literature search. Among

them, IGF2, ITGA5, C7, PLG, and TSC2 were also found in the Cancer Biomarker List [14] which is the data that was compared in the chapter 3.5. FBLN1 and FN1 were presented in the biomarker protein list that was provided by Clinical Proteomic Technology Assessment for Cancer (CPTAC) Program [30].

Table 6: The candidate biomarkers in early stage breast cancer found by pathway analysis and literature search

Gene	Annotation	Function	Notes	FC
IGF2	Insulin-like growth factor 2	Function: The insulin-like growth factors possess growth-promoting activity. In vitro, they are potent mitogens for cultured cells. IGF-II is influenced by placental lactogen and may play a role in fetal development.	A mitogenic growth factor; may have a role in fetal development. Increase in breast, prostate, lung and colorectum cancer.	1.08
IGFBP3	Insulin-like growth factor binding protein	Function: an insulin growth factor binding protein; involved in modulating IGF action.	Positive regulation of apoptosis, regulation of cell growth, positive regulation of myoblast differentiation, negative regulation of signal transduction. Increase in breast, prostate, lung and colorectum cancer.	1.14
SHC1*	src homology 2 domain-containing transforming protein C1	Key mediators of the insulin-like growth factor pathway, involved in transformation and differentiation in a Ras-dependent fashion.	Regulation of epidermal growth factor receptor activity; positive regulation of mitosis; positive regulation of cell proliferation and activation of MAPK activity. Actived in a high number of human tumors, including breast tumors.	1.22

FBLN1	fibulin 1	Function: Incorporated into fibronectin-containing matrix fibers. May play a role in cell adhesion and migration along protein fibers within the extracellular matrix (ECM). Could be important for certain developmental.	A secreted calcium-binding glycoprotein. Tumor suppressor. Altered expression of fibulin is associated with progression of several cancer types: bladder cancer breast cancer.	-1.06
FN1*	fibronectin 1	Function: Fibronectins bind cell surfaces and various compounds including collagen, fibrnin, heparin, DNA, and actin. Fibronectins are involved in cell adhesion, cell motility, opsoniztion, wound healing, and maintenance of cell shape.	Extracellular matrix component may play a role in fibrosis and anti tumor metastasis. It has been found to be regulated in prostate, thyroid and breast and ovarian cancer.	-1.23
ITGA5	integrin alph5	Function: Integrin alpha-5/beta-1 is a receptor for fibronectin and fibrinogen. It recognizes the sequence R-G-D in its ligands. In case of HIV-1 infection, the interaction with extracellular viral Tat protein seems to enhance angiogenesis in Kaposi's sar	Cell adhesion, integrin-mediated signaling pathway, cell-substrate junction assembly, alpha subunit that interacts with beta 1 subunit form a fibronectin receptor.	-1.14
C7,C5, C6 and C9	complement component	Function: C7 is a constituent of the membrane attack complex. C7 binds to C5b forming the C5b-7 complex, where it serves as a membrane anchor.	Component of membrane attack complex of complement, play a role in induction of apoptosis. Increase in lung cancer patient.	1.11

PLG	plasminogen	Function: Angiostatin is an angiogenesis inhibitor that blocks neovascularization and growth of experimental primary and metastatic tumors in vivo.	A tumor suppressor decreased in stomach and colonic cancer.	-1.06
ORM1*	Alpha-1-acid-glycoprotein1	Appears to function in modulating the activity of the immune system during the acute phase reaction and secreted	ORM and other acute-phase reactants may act as blocking factors protecting tumor cells against immunological attack, thereby contributing to the 'immune escape' of the tumor	1.36
TSC2	Isoform_1_of_Tuberin	Function: Implicated as a tumor suppressor. May have a function in vesicular transport, but may also play a role in the regulation of cell growth arrest and in the regulation of transcription mediated by steroid receptors. Interaction between TSC1 and TSC2 may facilitate vesicular docking. Specifically stimulates the intrinsic GTPase activity of the Ras-related protein RAP1A and RAB5. Suggesting a possible mechanism for its role in regulating cellular growth. A mutation in TSC2 leads to constitutive activation of RAP1A in tumors.	Regulation of progression through cell cycle and endocytosis.	1.12

*Proteins with fold change over 20%.

3.7 Results from subgroup analysis

Personalized medicine has become the next big wave toward improved drug development and patient care. In this study, the patient samples were very heterogeneous, including four different subgroups based on estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expressions. We therefore re-grouped them and re-analyzed the same data set. The subgroup information is shown in Table 7. Group P1 represents basal-like tumors type (ER-/PR-/HER2-, all negative), group P2 is HER2 positive (ER-/PR-/HER2+), Group P3 is luminal type A (ER+/PR+/HER2-), and group P4 is luminal type B (ER+/PR-/HER2). Four breast cancer subtypes were identified based on immunohistochemical (IHC) expression of the ER/PR and HER2 proteins [34].

Table 7: Experimental design for subgroup analysis

Group	Condition	Number of Samples
H	Healthy	40
P1	ER-/PR-/HER2-	10
P2	ER-/PR-/HER2+	8
P3	ER+/PR+/HER2-	6
P4	ER+/PR-/HER±	5

With the analysis of five different groups using 40 plasma samples from healthy controls and a total of 29 plasma samples from four different subtypes breast cancer patients (the other 11 patients samples were not used for analysis due to missing subtype information), we found most of the proteins that show differential expression are predominantly observed in Luminal type B (group P4) and Basal type (group P1). No significance changes with Her2+ (group P2) or luminal type A (group P3). All data are shown in Appendix 2 through Appendix 4, and summary results of group P1 and P4 are

shown in Table 8. Haptoglobin and S100A8 are two potential candidates suggested through this subgroup analysis shown in Table 9.

Table 8: Summary results of group P1 and P4

Proteins	Numbers	Priority	Significant Change	Max Absolute Fold. Change
P1 riched	34	1,2	Yes	2.1
P4 riched	54	1,2	Yes	2.1

Table 9: Biomarker candidates analyzed by subgrouping

Rank	Protein ID	Annotation	Max_FC	FC_H_P1	FC_H_P4	Function
1	IPI00641737	Haptoglobin _precursor	2.07	1.39	2.07	Haptoglobin is expressed by the liver and secreted in plasma. It combines with free plasma hemoglobin, preventing loss of iron through the kidneys and protecting the kidneys from damage by hemoglobin, while making the hemoglobin accessible to degradative enzymes.
28	IPI00007047	Protein_ S100_A8	1.43	1.01	1.43	Calcium binding protein. It is made in larger amountw in inflammatory diseases such as rheumatoid arthritis, and in some types of cancer.

With analysis of five different groups, a total of 1388 proteins were identified and quantified. The result from subgroup analysis is summarized in Table 10. Among the identified proteins, 471 were identified with high confidence (priority 1 and 2). Out of the 193 proteins that demonstrated significant changes, 77 were identified with high confidence. The highest fold change among priority 1 and priority 2 proteins was 2.6, and the overall maximum absolute fold change was 4.9 when comparing the differential expression between four sub-grouped patient samples to healthy controls.

Table 10: Overall summary of subgroup analysis

Protein Priority	Peptide ID Confidence	Multiple Sequences	Number of Proteins	Number Significant Changes	Max Absolute Foldchange	Median % CV Sample
1	High	Yes	226	46	2.2	15.9
2	High	No	246	31	2.6	24.8
3	Moderate	Yes	49	7	2.0	17.3
4	Moderate	No	867	109	4.9	22.6
Overall			1388	193	4.9	21.6

4. Discussion

It is believed that the nature of breast cancer is multi-factorial, a combination of several markers will add sensitivity and more specificity to effectively detect and diagnose cancer at its early stage. To look for such a panel of biomarker candidates of breast cancer, it requires not only high-throughput technologies, but also bioinformatics tools for complex data analysis. In this study, an LC/MS-based label-free method was applied to identify and quantify proteins in human plasma samples from patients with and without breast cancer. Through our analysis (Table 4), 1422 proteins were identified with high confidence, and 385 of them had significant changes between breast cancer patients and healthy controls. The many identified proteins provide a great opportunity to find novel biomarkers and predictive panels of biomarkers for early stage breast cancer. After pathway analysis and literature search (Table 6), 10 proteins were found to be related to a diseased state. However, among these 10 proteins, only 3 (FN1, ORM1, and SHC1) showed fold-changes greater than 20%, which is the cutoff for the study's median %CV. In this study, we considered the proteins whose fold-change was higher than its %CV because the fold-change is the ratio of the largest to the smallest protein intensities when comparing two groups and the coefficient of variation (CV) is a relative measure of sample variability. The %CV is the standard deviation divided by the mean on a % scale. The %CV may be interpreted as a % error. If a fold-change is less than a %CV change, then a fold-change may not represent a true biologically related change. In this study FN1 and ORM1 all are secretory proteins, and SHC1 is normally functioning within cells but released into plasma as a result of cell death or damage, which may serve as a panel of potential biomarkers.

As we know, early detection of breast cancer will improve the treatment and the survival rates of breast cancer patients because the ability to treat and cure breast cancer directly depends on the ability to detect it at its earliest stage. In this study, we used plasma samples with stages I and II breast cancer and healthy controls. After the analysis, we found a few proteins with significant fold-changes between the two groups, but the largest was only 53%. This is likely a result of low secretion of disease-related proteins into plasma during the early stages of breast cancer. A protein, or most likely a set of proteins, that undergoes changes in concentration or structural composition (e.g., PTM) as a result of disease or physiological state may occur more often and with greater extent during the later stage of cancer progression [16, 18]. The important proteins (such as a secretory protein FBLN1) in the signaling pathway, may have very low fold-change in this study, and can be misinterpreted.

The results from subgroup analysis show that all proteins differentially expressed between healthy subjects and those who have breast cancer belong to either Basal type-triple-negative (P1 with worst prognosis) or Luminal type B (P4), but not Her 2+ (P2) or Luminal type A (P3) (Table 7). The fold changes among the subgroups were significantly higher than before sub-grouping. This data is consistent with recent literature, which suggests that many malignancies, including breast cancer, can be more effectively treated by first dividing the population into subgroups [35]. Pharmaceutical companies use these markers from sub-grouping to develop new drugs more efficiently [36]. The drugs are not only more successful at treating patients, but responses to these drugs are more predictable as well [36]. For example, the expressions of ER, PR and HER2 have been used clinically as predictors of trastuzumab response and hormonal

treatment for breast cancer patients [35]. Hopefully, the results from this subgroup analysis can provide useful biomarker candidates to better identify drug targets to effectively treat certain types of breast cancer.

4.1 The role of the IGF system in cancer growth and metastasis (SHC1)

The possibility of the Src homology 2 domain-containing transforming protein C1 (SHC1) being a breast cancer biomarker candidate is supported by numerous other studies. SHC1 is a polymorphism that is located in important regulatory regions, such as in the promoter region of functionally active domains, which may affect the function of protein [37]. Insulin receptor substrates (IRS) and Src homology 2 domains—containing transforming protein 1's (SHC1)—are key mediators of the insulin-like growth factor pathway [40, 41]. They act as docking proteins between the activated receptors and the further downstream signaling proteins. IRS is a major substrate for the IGF receptor while the iRS2 and SHC1 proteins are tyrosine phosphorylated and can therefore interact and activate Scr homology 2 (SH2) domain containing proteins. Both IRS and SHC1 can bind to growth factor receptor bound protein2 (GRB2) to activate the Ras/MAP kinase pathway that regulates cell proliferation. Changes in the activity of the IRS and SHC1 proteins may thus have an effect both on cell proliferation and apoptosis. It has been shown that SHC1 is constitutively activated in a high number of human tumors, including breast tumors [42], and may explain the over expression of MAP kinase detected in many cancers [42, 43]. Since the fold change (1.22) is higher than the %CV (20%) of this protein in this study, there is further evidence to suggest that the SHC1 is expressed in different quantities in healthy and cancer patients. Figure 4.1.1 shows the relative

expression levels of Src homology 2 domain-containing transforming protein C1 when comparing cancer sample group with control sample group.

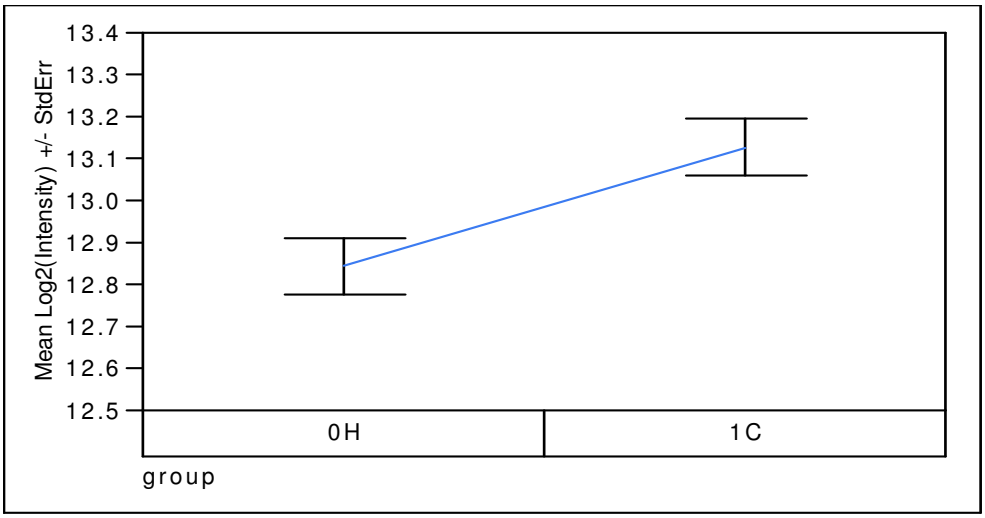

Figure 4.1.1 Gene Name: SHC1, Protein Id=IPI00021326.4, SigChange=Yes, Annotation=SHC_ (Src_homology_2_domain_containing) _transforming_protein_1. Relative expression levels of SHC1 under different conditions. The intensities which are given by the AUC from the XIC are transformed to the log base 2 scale, base 2 is popular because a two-fold change is transformed to a unit change on a log base 2 scale. Error bars show standard errors based on the ANOVA model.

4.2 Control of cell motility and tumor invasion by extracellular matrix interactions (FN1)

Another reasonable breast cancer biomarker candidate in this study is the fibronectin1 protein (FN1). Fibronectin (FN) is an example of an extracellular matrix (ECM) protein that promotes cell motility [46]. Cell movements largely depends on cell interactions with specific components of the ECM and is a fundamental feature of many normal and pathological processes, including developmental morphogenesis,

inflammation, wound healing, and metastasis [45]. Fibulin-1 (FBIN1), a secreted calcium-binding glycoprotein assembled as insoluble polymers and presented in the blood as a soluble dimer, is found in association with ECM structures such as micro fibrils, basement membranes, elastic fibres, and fibrin. The association of fibulin-1 with these ECM structures is probably based on its ability to bind ECM proteins such as fibronectin (FN), laminin, nidogen, and fibrinogen [47]. FN1 is a homologue of FN and contains a self-assembly domain, which induces FN1-FN1 polymerization [49]. FBLN1 has been implicated in cellular transformation and tumor invasion. It appears to be a tumor suppressor and may play a role in homeostasis and thrombosis owing to its ability to bind fibrinogen and incorporate it into clots [48]. Strong support for in vivo anti-tumor activity of polymeric FN1 exhibited anti-tumor activity in mice bearing various types of tumors [50]. Another report [51] showed that anastellin, a component of FN1 that is capable of inducing FN1 polymerization, displayed anti-angiogenic and anti-metastastic properties. These findings provide a rational explanation for this study. Figure 4.2.1 shows the relative expression levels of fibronectin 1 protein when comparing cancer sample group with control sample group, the fold change is -1.23.

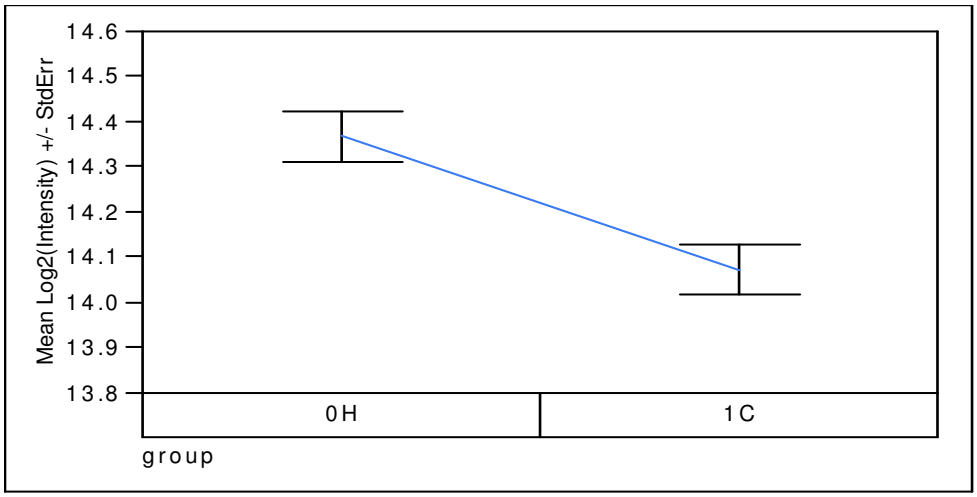

Figure 4.2.1 Relative expression levels of FN1 under different conditions

Gene Name=FN1, variability Gauge Rank=17, Protein Id=IPI00411462.4, SigChange=Yes, Annotation=Fibronectin_1

4.3 The role of acute-phase reactants in the modulation of the immune system: protein Orosomucoid1 (ORM1)

Orosomucoid1 (ORM1) was also found to be differentially expressed in healthy and cancer patient in this study—higher levels in early stage breast cancer patients. A Standard Error chart for this protein is shown in Figure 4.3.1. Alpha1-acid glycoprotein, first described in 1950 by Schmid [38], belongs to a group of acute-phase proteins that may play a role in the modulation of the immune system in response to stress, along with many other functions [38]. In literature there are several studies concerning ORM1 glycoform determination for the diagnosis of different types of cancer, such as colorectal [39] and ovarian cancer [39]. Also, serum level of ORM is increased in inflammatory and lymphoproliferative disorders and cancers, such as breast, lung and ovarian cancers [38], leading to the assumption that ORM might be produced by cancer cells themselves [39]. Therefore, ORM1 is an important candidate to study to increase the probability of

early detection in cancer patients. The differential expression level of this protein (the fold change is 1.36) in different conditions is shown in Figure 4.3.1 in this study.

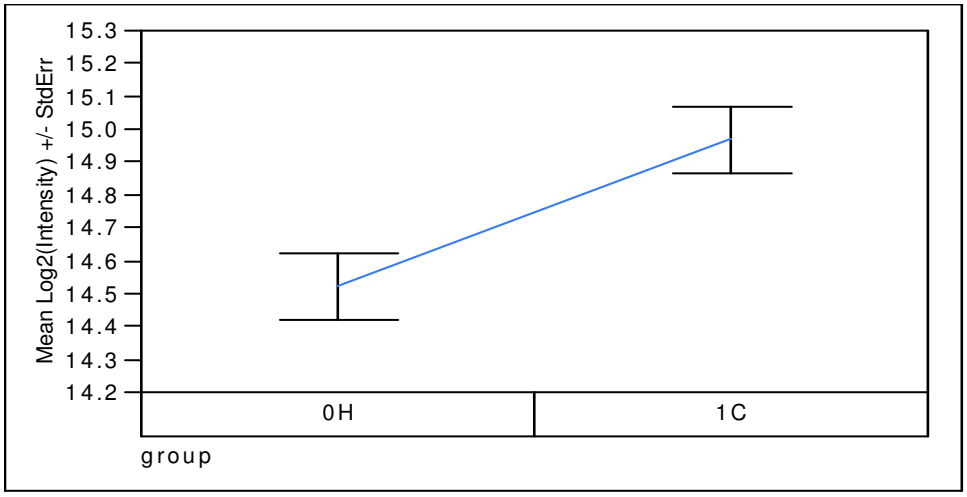

Figure 4.3.1 Relative expression levels of ORM1 under different conditions

Gene Name=ORM1, variability Gauge Rank=37, Protein Id=IPI00022429.3, SigChange=Yes, Annotation=Alpha-1-acid_glycoprotein_1_precursor

4.4 Proteins from subgrouping analysis: Haptoglobin (HP) and S100A8 as potential biomarkers for Luminal type B and Basal type breast cancer

One of the major proteins, Haptoglobin, in subgroup P1 and P4 is expressed by the liver and secreted in plasma. It combines with free plasma hemoglobin, preventing loss of iron through the kidneys and protecting the kidneys from damage by hemoglobin, while making the hemoglobin accessible to degradative enzymes. Haptoglobin has recently been demonstrated as a serum marker for prostate cancer [44] and ovarian cancer [52]. In this study, haptoglobin is the strongest biomarker candidate increased in Luminal type B (2.1-fold) but less so in Basal (1.4-fold). This protein has previously been shown to be up-regulated in breast cancer and is associated with metastasis

recurrence [53]. It is up-regulated by Nuclear factor-KB (NF-KB is a ubiquitous transcription factor that controls the expression of genes involved in immune responses, apoptosis, and cell cycle), but this induction is repressed by estrogen receptor. The loss of estrogen receptor and NF-KB crosstalk may lead to elevated expression [53].

The next candidate S100 A8 is a calcium binding protein. It is made in larger amounts in inflammatory diseases such as rheumatoid arthritis, and in some types of cancer [54]. In this study, it increased in Luminal B (1.5-fold) but not basal (1-fold). It induces NF-KB activity [55]. Previous studies have shown its elevation in basal-type breast cancer [54]. In our studies it is elevated in Luminal type B. This protein has been shown in a few literatures to be related with other cancers, for example bladder cancer [56]. The differential expression levels of haptoglobin and S100 A8 in the subgroups are shown Figure 4.4.1 and Figure 4.4.2 below.

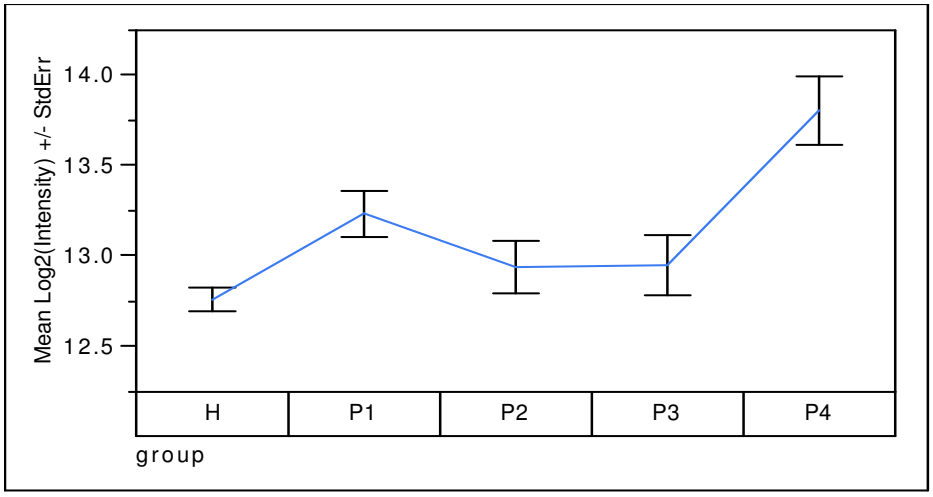

Figure 4.4.1 Relative expression levels of haptoglobin under different conditions

Rank=1, Protein Id=IPI00641737.1, SigChange=Yes, Annotation=Haptoglobin

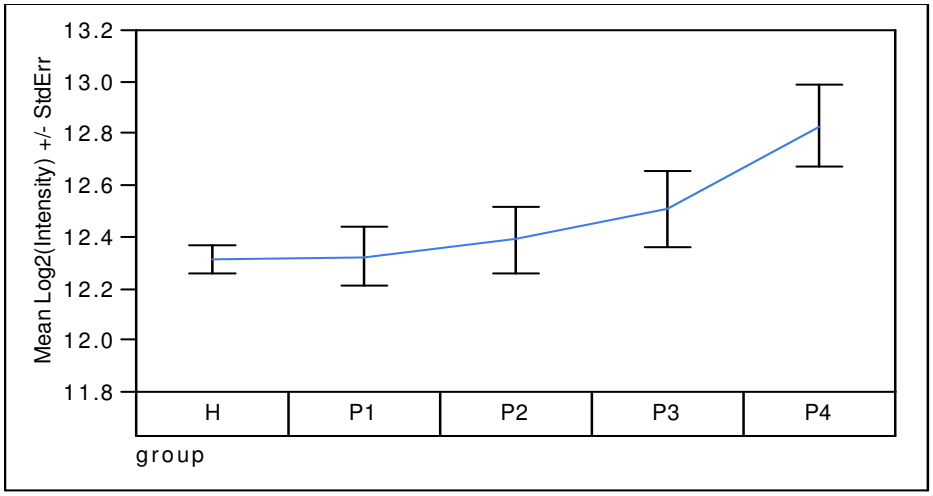

Figure 4.4.2 Relative expression levels of S100 A8 under different conditions

Rank=28, Protein Id=IPI00007047.1, SigChange=Yes, Annotation=Protein_S100-A8

4.5 Summary

From this study, Src homology 2 domain-containing transforming protein C1 (SHC1), fibronectin 1 protein (FN1) and orosomucoid 1 protein (ORM1) were selected as biomarker candidates in early stage breast cancer when comparing 40 breast cancer patients and healthy control samples. Haptoglobin protein (HP) and S100A8 protein

were selected as biomarker candidates for Luminal type B and Basal type breast cancer when the breast cancer patients were divided into four sub-groups and compared with healthy control samples. Sub-grouping can lead to a more targeted approach personalized medicine in the development of breast cancer treatments, tailoring treatments to a specific type of breast cancer. The fold changes after sub-grouping were significantly higher than normal grouping. The data from sub-grouping was consistent with recent literature in which pharmaceutical companies used biomarkers from sub-grouping to develop personalized medicine more efficiently [36]. As a follow up, a study to validate these selected biomarker candidates with and without sub-grouping is necessary.

5. Conclusion

The results from this study showed that LC/MS label-free protein quantification technology with statistical analysis allows for detection of, not only more low-abundance proteins, but also proteins with small changes in expression. This method along with bioinformatics analysis provides an excellent opportunity to identify potential biomarkers for early stage breast cancer, however, any discoveries made by proteomic approaches should be validated using different methods. Once these potential biomarkers are confirmed and validated, they can then serve as targets for early diagnostics and therapeutics development.

Appendix 1: The information of all 385 proteins (priority 1-4) with significant changes (q value less than 5%) using LC/MS-based label free protein quantification method

Protein_ID	Gene symbol	FC	qValue	Annotation	Biological Process
IPI00022431.1	AHSG	-1.20	0.001	Alpha-2-HS-glycoprotein_precursor	P: regulation of inflammatory response P: positive regulation of phagocytosis P: pinocytosis P: negative regulation of insulin receptor sig... P:negative regulation of bone mineralization P:acute-phase response
225768	SAA1	1.21	0.001	1313184B_alpha1_antitrypsin	P:electron transport
IPI00164623.4	C3	1.12	0.001	187_kDa_protein	
IPI00783987.1	C3	1.12	0.001	Complement_C3_precursor_(Fragment)	P: immune response P: G-protein coupled receptor protein signalin...
671882	APOA2	-1.28	0.001	apolipoprotein_[Homo_sapiens]	P: response to glucose stimulus P: regulation of cytokine production P: positive regulation of interleukin-8 biosyn... P: neutrophil activation P: negative regulation of lipoprotein metaboli... P:negative regulation of lipid catabolic process P:lipid transport P:glucose metabolic process
IPI00021854.1	APOA2	-1.30	0.002	Apolipoprotein_A-II_precursor	
IPI00305457.5	SERPINA1	1.18	0.002	PRO2275	
IPI00431656.4	SERPINA3	1.15	0.002	Isoform_2_of_Alpha-1-antichymotrypsin_precurs	P:regulation of lipid metabolic process P:inflammatory response
223433	DAP13	1.17	0.002	0805684B_inhibitor,alpha1_protease	
225769		1.13	0.002	1313184C_chymotrypsin_inhibitor	
IPI00550991.2	SERPINA3	1.13	0.003	Isoform_1_of_Alpha-1-antichymotrypsin_precursor	
IPI00294004.1	PROS1	1.09	0.004	Vitamin_K-dependent_protein_S_precursor	P:blood coagulation
IPI00553177.1	SERPINA1	1.16	0.004	Alpha-1-antitrypsin_precursor	
40737516	C4B	1.21	0.004	C4A6_[Homo_sapiens]	
IPI00032258.4	C4A	1.18	0.004	Complement_C4-A_precursor	

177836	SERP1NA1	1.15	0.004	alpha-1-antitrypsin_precursor	
IPI00411462.4	FN1	-1.23	0.004	fibronectin_1_isoform_7_preproprotein	
IPI00418163.3	C4B	1.17	0.004	complement_component_4B_preproprotein	
40737408	C4B	1.17	0.004	C4B5_[Homo_sapiens]	P:complement activation
IPI00003351.2	ECM1	-1.09	0.006	Extracellular_matrix_protein_1_precursor	P: positive regulation of I-kappaB kinase/NF-k...
229453		-1.09	0.006	731316A_plasminogen_1-69	
IPI00028413.7	ITIH3	1.11	0.007	Inter-alpha-trypsin_inhibitor_heavy_chain_H3_precursor	
4758236	ECM1	-1.08	0.008	extracellular_matrix_protein_1_isoform_1_precursor_[Homo_sapiens]	
3024064	ITIH3	1.11	0.009	ITIH3_HUMAN_Inter-alpha-trypsin_inhibitor_heavy_chain_H3_precursor_(ITI_heavy_chain_H3)_(Inter-alpha-inhibitor_heavy_chain_3)_(Serum-derived_hyaluronan-associated_protein)_(SHAP)	
30722344		-1.19	0.009	hypothetical_protein_[Homo_sapiens]	P: transmembrane receptor protein tyrosine kin... P:response to wounding P:cell migration P:cell adhesion
IPI00022418.1	FN1	-1.19	0.009	Isoform_1_of_Fibronectin_precursor	
IPI00022395.1	C9	1.11	0.009	Complement_component_C9_precursor	P:hemolysis by symbiont of host red blood cells
IPI00296165.5	C1R	1.09	0.011	Complement_C1r_subcomponent_precursor	P:immune response
34365361	DKFZp686O12165	-1.18	0.012	hypothetical_protein_[Homo_sapiens]	P:acute-phase response
223962	C4d	1.15	0.013	1006226B_complement_C4d_variant	P: carbon utilization by fixation of carbon di...
IPI00019580.1	PLG	-1.06	0.014	Plasminogen_precursor	P: negative regulation of cell proliferation P:negative regulation of blood vessel endothe... P:negative regulation of angiogenesis P:induction of apoptosis P:blood coagulation
IPI00009920.2	C6	1.07	0.014	Complement_component_C6_precursor	

13787109		1.13	0.014	A_Chain_A,_A_2.1_Angstrom_Structure_Of_An_Uncleaved_Alpha-1-_Antitrypsin_Shows_Variability_Of_The_Reactive_Center_And_Other_Loops	
IPI00032291.1	C5	1.11	0.014	Complement_C5_precursor	P: inflammatory response P: G-protein coupled receptor protein signalin... P:chemotaxis P:activation of MAPK activity
899271	C7	1.11	0.015	complement_C7_[Homo_sapiens]	P:complement activation
IPI00017696.1	C1S	1.07	0.015	Complement_C1s_subcomponent_precursor	
IPI00022429.3	ORM1	1.36	0.015	Alpha-1-acid_glycoprotein_1_precursor	P:acute-phase response
38488662	C6	1.07	0.015	complement_component_6_[Homo_sapiens]	P:complement activation
229386		1.43	0.017	720005A_protein,alpha1_acid_glyco	
IPI00019943.1	AFM	-1.11	0.017	Afamin_precursor	
IPI00303152.4	COL22A1	-1.09	0.017	collagen,_type_XXII,_alpha_1	P:cell adhesion
21707947	ECM1	1.15	0.017	Leucine-rich_alpha-2-glycoprotein_1_[Homo_sapiens]	
IPI00001611.1	IGF2	1.08	0.020	Isoform_1_of_Insulin-like_growth_factor_II_precursor	P:skeletal development P:insulin receptor signaling pathway P:genetic imprinting
39725934	SERPINF1	1.08	0.023	serine_(or_cysteine)_proteinase_inhibitor,_clade_F_(alpha-2_antiplasmin,_pigment_epithelium_derived_factor),_member_1_[Homo_sapiens]	P:positive regulation of neurogenesis P:negative regulation of angiogenesis P:cell proliferation
179716	C7	1.11	0.023	complement_protein_C7_precursor	
IPI00009793.2	C1RL	1.05	0.023	Complement_component_1,_r_subcomponent-like_variant_(Fragment)	P:proteolysis
IPI00296608.6	C7	1.11	0.025	Complement_component_C7_precursor	
IPI00006114.4	SERPINF1	1.08	0.025	Pigment_epithelium-derived_factor_precursor	
IPI00022426.1	AMBP	1.08	0.026	AMBP_protein_precursor	P:negative regulation of JNK cascade P:negative regulation of immune response P:heme catabolic process P:female pregnancy P:cell adhesion

IPI00027462.1	S100A9	1.10	0.026	Protein_S100-A9	P:inflammatory response P:cell-cell signaling
IPI00026199.1	GPX3	-1.08	0.026	Glutathione_peroxidase_3_precursor	P:response to lipid hydroperoxide P:protein homotetramerization P:hydrogen peroxide catabolic process
13937839	SAA1	1.14	0.029	SAA1_protein_[Homo_sapiens]	P: positive regulation of interleukin-1 secretion P: positive regulation of cell adhesion P: platelet activation P:neutrophil chemotaxis P:negative regulation of inflammatory response P:macrophage chemotaxis P:lymphocyte chemotaxis P:elevation of cytosolic calcium ion concentr... P:acute-phase response
10436374		1.05	0.037	unnamed_protein_product_[Homo_sapiens]	P:proteolysis
178818	APOB	-1.08	0.038	apolipoprotein_B-100	P:signal transduction P:lipid transport P:blood circulation
40316910	SAA2	1.16	0.040	serum_amyloid_A1_isoform_2_[Homo_sapiens]	
15029894	SERPING1	1.08	0.040	Serpin_peptidase_inhibitor,_clade_G_(C1_inhibitor),_member_1,_(angioedema,_hereditary)_[Homo_sapiens]	P:blood circulation
IPI00022417.4	LRG1	1.16	0.040	Leucine-rich_alpha-2-glycoprotein_precursor	
229479		-1.19	0.046	740525A_lipoprotein_Gln_I	
IPI00002286.5	ANKRD11	1.23	0.000	Ankyrin_repeat_domain-containing_protein_11	
IPI00217617.4	MPP7	-1.20	0.000	palmitoylated_membrane_protein_7	
IPI00218852.2	VIL1	-1.24	0.000	Villin-1	P:protein complex assembly P:actin filament severing
IPI00748360.1	KIAA1797	-1.30	0.000	Novel_gene	
IPI00295640.3	ABCA6	-1.30	0.000	Isoform_1_of_ATP-binding_cassette_sub-family_A_member_6	
IPI00011245.1	USP29	1.20	0.000	Ubiquitin_carboxyl-terminal_hydrolase_29	P:protein modification process
IPI00445774.1	GPR113	-1.28	0.000	CDNA_FLJ43404_fis,_clone_OCBBF2017516	
IPI00165955.1	MAPK15	1.16	0.001	Isoform_1_of_Mitogen-activated_protein_kinase_15	P:protein amino acid autophosphorylation
IPI00375286.3	KIAA0802	-1.19	0.001	KIAA0802_protein	
IPI00333913.7	NAG	-1.45	0.001	neuroblastoma-amplified_protein	

IPI00026108.1	PLA2G4A	1.17	0.001	Cytosolic_phospholipase_A2	P:platelet activating factor biosynthetic process P:icosanoid metabolic process
IPI00440493.2	ATP5A1	1.18	0.001	ATP_synthase_subunit_alpha,_mitochondrial_precursor	
IPI00003843.1	TJP2	1.13	0.002	Isoform_A1_of_Tight_junction_protein_ZO-2	
IPI00465430.5		-1.26	0.002	70_kDa_protein	
IPI00219561.1	NLRP14	1.18	0.002	NACHT,_LRR_and_PYD-containing_protein_14	P:spermatogenesis
IPI00001364.2	C21orf66	1.18	0.002	Isoform_A_of_GC-rich_sequence_DNA-binding_factor_homolog	P:regulation of transcription, DNA-dependent
IPI00025418.1	COL7A1	-1.28	0.002	Collagen_alpha-1(VII)_chain_precursor	P:epidermis development
IPI00152774.1	ARL6IP4	1.16	0.003	FLJ00169_protein_(Fragment)	
IPI00256429.3	MYO7B	1.19	0.004	PREDICTED:_myosin_VIIB	
IPI00009891.1	TAF1	-1.20	0.004	Isoform_1_of_Transcription_initiation_factor_TFIID_subunit_1	P:protein amino acid autophosphorylation P:G1 phase of mitotic cell cycle
IPI00045550.3	PPP1R9B	1.22	0.004	Neurabin-2	
32967319		1.14	0.005	ephrin_receptor_EphA5_isoform_b_[Homo_sapiens]	
8885790	FLNC	-1.24	0.005	AF146692_1_filamin_2_[Homo_sapiens]	
IPI00029449.6	VANGL2	1.14	0.005	Vang-like_protein_2	P:sensory cilium biogenesis P:neural tube closure P:heart looping P:establishment of planar polarity P:apical protein localization
IPI00413728.3	SPTAN1	1.14	0.005	Isoform_1_of_Spectrin_alpha_chain,_brain	
IPI00100460.2	DARS2	-1.34	0.005	Aspartyl-tRNA_synthetase,_mitochondrial_precursor	
IPI00304267.3	NXN	1.23	0.007	nucleoredoxin	P:electron transport
IPI00412404.1	SUPV3L1	-1.32	0.007	Putative_ATP-dependent_mitochondrial_RNA_helicase	
IPI00303944.6	NSUN4	1.08	0.007	NOL1/NOP2/Sun_domain_family_4_protein	
IPI00003590.2	QSOX1	1.13	0.009	Isoform_1_of_Sulfhydryl_oxidase_1_precursor	P:protein thiol-disulfide exchange
IPI00014215.5	HYDIN	1.16	0.009	HYDIN_protein	
IPI00169256.1	HBE269	-1.12	0.011	HBE269	
IPI00303530.4	MTHFSD	1.09	0.011	CDNA_FLJ13893_fis,_clone_THYRO1001661	

IPI00022470.2	ZN516	1.17	0.011	PREDICTED:_similar_to_Zinc_finger_protein_516	
IPI00030059.1	GNG10	-1.15	0.011	Guanine_nucleotide-binding_protein_G(I)/G(S)/G(O)_gamma-10_subunit_precursor	P:signal transduction
IPI00293887.4	STARD8	-1.12	0.011	Hypothetical_protein_DKFZp686H1668	P:signal transduction
IPI00550021.3	RPL3	1.14	0.011	60S_ribosomal_protein_L3	P:translation
IPI00413895.2	GOLGA2	-1.21	0.011	Golgi_autoantigen,_golgin_subfamily_A_member_2	
IPI00102936.3	SRP68	-1.21	0.011	Isoform_2_of_Signal_recognition_particle_68_kDa_protein	
IPI00003515.1	TRIP11	1.21	0.011	Thyroid_receptor-interacting_protein_11	P:transcription from RNA polymerase II promoter
IPI00002564.3	XRCC1	1.22	0.011	DNA-repair_protein_XRCC1	
IPI00029661.2	EVI5	1.22	0.011	ecotropic_viral_integration_site_5	P:multicellular organismal development P:cell proliferation
IPI00465166.1	ATP8A2	1.11	0.012	CDNA_FLJ45330_fis,_clone_BRHIP3007195,_highly_similar_to_Potential_phospholipid-transporting_ATPase_IB	P:cation transport
IPI00008456.1	HSF4	1.11	0.013	Isoform_HSF4B_of_Heat_shock_factor_protein_4	P:regulation of transcription, DNA-dependent
IPI00332837.1	PLEKHA7	-1.21	0.014	PLEKHA7_protein_(Fragment)	
IPI00166085.1	LGI2	-1.26	0.016	Leucine-rich_repeat_LGI_family_member_2_precursor	
IPI00096899.6	CDT1	-1.16	0.016	DNA_replication_factor_Cdt1	P:DNA replication checkpoint
IPI00165984.4	WDR35	1.15	0.016	Isoform_1_of_WD_repeat_protein_35	
IPI00175649.3	LRRK2	1.08	0.016	Leucine-rich_repeat_serine/threonine-protein_kinase_2	
IPI00060148.3	CCDC127	-1.10	0.016	Coiled-coil_domain_containing_127	
IPI00744612.1	PRDM5	1.22	0.016	Conserved_hypothetical_protein	
IPI00018903.1	VAX2	1.36	0.016	Ventral_anterior_homeobox_2	P:visual perception P:ectoderm development
IPI00556155.1	IGFBP3	1.14	0.016	insulin-like_growth_factor_binding_protein_3_isoform_a_precursor	

IPI00156532.1	ULK4	1.19	0.017	Hypothetical_protein_FLJ20574	
IPI00244116.1	C1ORF67	-1.12	0.019	Novel_protein	
IPI00155475.5	KCTD19	-1.18	0.021	KCTD19_protein	
IPI00293925.2	FCN3	1.13	0.021	Isoform_1_of_Ficolin-3_precursor	
IPI00013492.1	TCF4	-1.18	0.021	Isoform_SEF2-1B_of_Transcription_factor_4	P: positive regulation of transcription, DNA-d...
IPI00218803.2	FBLN1	-1.06	0.024	Isoform_B_of_Fibulin-1_precursor	
IPI00298935.4	JMJD1B	-1.18	0.025	Isoform_1_of_JmjC_domain-containing_histone_demethylation_protein_2B	
8176715	GIPR	1.08	0.026	glucose-dependent_insulinotropic_polypeptide_receptor;_GIP_receptor_[Homo_sapiens]	P:response to nutrient P:generation of precursor metabolites and energy P:adenylate cyclase activation
IPI00787774.1	LOC730756	-1.22	0.026	PREDICTED:_similar_to_collagen,_type_I,_alpha_1	
IPI00736556.3	KIAA1486	-1.12	0.026	PREDICTED:_similar_to_myosin_heavy_chain_Myr_8_isoform_1	
IPI00005794.1	PGCP	-1.12	0.026	Blood_plasma_glutamate_carboxypeptidase_precursor	
22760630	WDR31	-1.17	0.027	unnamed_protein_product_[Homo_sapiens]	
IPI00069084.1	TRRAP	1.10	0.028	Isoform_1_of_Transformation/transcription_domain-associated_protein	P:signal transduction P:histone acetylation
IPI00385791.1	MRPS26	1.17	0.031	Serologically_defined_breast_cancer_antigen_NY-BR-87_(Fragment)	
30268327	TXLNA	-1.23	0.032	hypothetical_protein_[Homo_sapiens]	P:cell proliferation
IPI00026940.2	NUP50	-1.16	0.032	Nucleoporin_50_kDa	
IPI00549941.2	LOC220115	1.23	0.033	TPTEps1_protein_(Fragment)	
IPI00761046.1		1.10	0.036	Conserved_hypothetical_protein	
IPI00008556.1	F11	1.06	0.036	Isoform_1_of_Coagulation_factor_XI_precursor	P:blood coagulation
IPI00215631.1	VCAN	-1.25	0.036	Isoform_Vint_of_Versican_core_protein_precursor	P:multicellular organismal development P:cell recognition P:cell adhesion
IPI00339217.2	OVCH1	1.17	0.041	ovochymase_1	
IPI00028493.3	TSC2	1.12	0.042	Isoform_1_of_Tuberin	P:regulation of progression through cell cycle P:endocytosis

IPI00182757.9	KIAA1967	1.09	0.047	Isoform_2_of_Protein_KIAA1967	
IPI00177967.6	MPHOSPH9	1.16	0.001	M-phase_phosphoprotein_9	
IPI00179164.2	KIAA1244	-1.19	0.002	hypothetical_protein_LOC57221	
IPI00299503.1	GPLD1	-1.10	0.004	Phosphatidylinositol-glycan-specific_phospholipase_D_1_precursor	
IPI00023824.1	FBLN2	1.10	0.006	Fibulin-2_precursor	
IPI00397927.1	FLJ46347	1.14	0.009	CDNA_FLJ46347_fis,_clone_TESTI4047437	
IPI00398808.5	WNK2	-1.08	0.023	Isoform_1_of_Serine/threonine-protein_kinase_WNK2	P:protein kinase cascade P:protein amino acid phosphorylation
IPI00103516.1	CASKIN2	-1.08	0.023	Caskin-2	
IPI00152462.1	DNAH3	1.05	0.027	Axonemal_heavy_chain_dynein_type_3	
IPI00299635.3	BIRC6	-1.09	0.027	Baculoviral_IAP_repeat-containing_protein_6	P:anti-apoptosis
IPI00218539.3	COL11A1	-1.09	0.028	Isoform_B_of_Collagen_alpha-1(XI)_chain_precursor	P: visual perception P: extracellular matrix organization and bioge... P:cartilage condensation
IPI00026320.1	UBR5	1.08	0.030	Ubiquitin-protein_ligase_EDD1	P:ubiquitin-dependent protein catabolic process P:response to DNA damage stimulus P:progesterone receptor signaling pathway P:cell proliferation
IPI00411743.5	CCDC15	1.09	0.033	CCDC15_protein_(Fragment)	
IPI00102752.2	RBM15	1.09	0.038	Isoform_1_of_Putative_RNA-binding_protein_15	
IPI00289275.2	CILP	-1.31	0.001	Cartilage_intermediate_layer_protein_1_precursor	P: nucleobase, nucleoside, nucleotide and nucl...
IPI00031666.4	RP3-402G11.5	-1.29	0.001	Selenoprotein_O	
IPI00026321.3	GFM1	-1.37	0.001	CDNA_FLJ12662_fis,_clone_NT2RM4002205,_moderately_similar_to_ELONGATION_FACTOR_G,_MITOCHONDRIAL	
IPI00749092.1		-1.30	0.001	Similar_to_Phosphoinositide_3-phosphatase	
IPI00455122.1		-1.30	0.001	Novel_protein	
IPI00219005.2	FKBP4	-1.30	0.001	FK506-binding_protein_4	P:protein folding
1097307		-1.26	0.001	2113367A_HIC-1_gene	P:metabolic process
IPI00376259.1	SRGAP1	1.17	0.001	Isoform_1_of_SLIT-ROBO_Rho_GTPase-activating_protein_1	

IPI00250364.1	TAS2R60	-1.50	0.001	Taste_receptor_type_2_member_60	P:sensory perception of bitter taste
IPI00011875.1	IK	-1.49	0.001	Protein_Red	P:immune response P:cell-cell signaling
3024656		1.18	0.001	SOX12_HUMAN_SOX-12_protein_(SOX-22_protein)	P: regulation of transcription from RNA polyme...
IPI00477676.2	LOXHD1	1.18	0.001	LOXHD1_protein	
IPI00102677.1	TESK2	1.17	0.001	Isoform_1_of_Dual_specificity_testis-specific_protein_kinase_2	P:spermatogenesis P:protein amino acid phosphorylation P:focal adhesion formation P:actin cytoskeleton organization and biogenesis
IPI00455063.4	MIXED	1.22	0.001	PREDICTED:_similar_to_leiomodin_2	
IPI00746871.1		1.18	0.001	Similar_to_Alu_subfamily_SQ_sequence_contamination_warning_entry	
IPI00029501.1	SEBOX	1.20	0.001	SEBOX	
IPI00651656.1	GRASP	1.20	0.001	Isoform_2_of_General_receptor_for_phosphoinositides_1-associated_scaffold_protein	
IPI00166215.4	INTS1	-1.16	0.001	Integrator_complex_subunit_1	
IPI00167515.1	ZADH1	1.18	0.001	Isoform_1_of_Zinc-binding_alcohol_dehydrogenase_domain-containing_protein_1	
IPI00291186.5	SLC7A13	1.18	0.001	solute_carrier_family_7,_(cationic_amino_acid_transporter,_y+_system)_member_13	P:amino acid transport
IPI00004939.1		1.20	0.001	Neuronal_thread_protein_AD7c-NTP	P:central nervous system development P:apoptosis
IPI00216662.5	TMPRSS6	1.20	0.001	Isoform_3_of_Transmembrane_protease,_serine_6	P: proteolysis P: intracellular signaling cascade P: fibrinolysis P:extracellular matrix organization and bioge... P:angiogenesis
IPI00295368.1	GCNT1	1.18	0.001	Beta-1,3-galactosyl-O-glycosyl-glycoprotein_beta-1,6-N-_acetylglucosaminyltransferase	P:protein amino acid O-linked glycosylation
IPI00002459.4	ANXA6	1.18	0.001	annexin_VI_isoform_2	
IPI00102291.1	FAM126A	1.18	0.001	Protein_FAM126A	
IPI00783147.1	LOC401387	1.18	0.001	84_kDa_protein	
IPI00011547.2	NAIP	-1.30	0.001	Baculoviral_IAP_repeat-containing_protein_1	P:nervous system development P:anti-apoptosis

IPI00012842.1	SLC26A4	-1.15	0.001	Pendrin	P:sulfate transport P:sensory perception of sound
IPI00787902.1		-1.14	0.001	PREDICTED:_similar_to_Serine_protease_inhibitor_Kazal-type_7_precursor	
IPI00293593.2	TCTE1	1.18	0.001	OTTHUMP00000016524	
IPI00021388.1	MED13	1.18	0.001	Thyroid_hormone_receptor-associated_protein_complex_240_kDa_component	P: transcription initiation from RNA polymeras... P: positive regulation of transcription from R... P:androgen receptor signaling pathway
IPI00409601.1	KIAA1219	-1.34	0.001	Isoform_1_of_Protein_KIAA1219	
IPI00394796.4	TDRD1	1.18	0.001	tudor_domain_containing_1	
IPI00169306.1	PFN4	-1.14	0.001	Profilin-4	
IPI00028064.1	CTSG	1.15	0.001	Cathepsin_G_precursor	P:proteolysis P:immune response
IPI00399257.2	DCHS2	1.16	0.001	CDNA_FLJ45804_fis,_clone_NT2RI3005923,_weakly_similar_to_Cadherin-_related_tumor_suppressor	P:homophilic cell adhesion
IPI00023258.5	TAF4B	1.17	0.001	PREDICTED:_similar_to_Transcription_initiation_factor_TFIID_105_kDa_subunit	
IPI00217972.4	COL27A1	-1.18	0.001	Collagen_XXVII_proalpha_1_chain_precursor	
IPI00410330.3	GATAD2A	-1.26	0.002	Isoform_1_of_Transcriptional_repressor_p66_alpha	P: negative regulation of transcription, DNA-d... P:DNA methylation
IPI00029928.2	ELN	1.17	0.002	Isoform_2_of_Elastin_precursor	P:respiratory gaseous exchange P:organ morphogenesis P:blood circulation P:cell proliferation
33187736		1.19	0.002	PCPD_protein_[Homo_sapiens]	
IPI00145121.5	PIGB	-1.15	0.002	GPI_mannosyltransferase_3	P:GPI anchor biosynthetic process
IPI00005822.1	CDC23	-1.25	0.002	Isoform_1_of_Cell_division_cycle_protein_23_homolog	P:cell division
IPI00006602.3	NDEL1	-1.17	0.002	nudE_nuclear_distribution_gene_E_homolog_like_1_(A._nidulans)_isoform_A	
IPI00400922.4	PDCD11	1.16	0.003	RRP5_protein_homolog	
IPI00387132.5	hCG_2007354	-1.27	0.003	CDNA_FLJ32661_fis,_clone_TESTI1000055,_weakly_similar_to_HOMEOBOX_PROTEIN_SIX1	P:regulation of transcription, DNA-dependent

IPI00414320.1	ANXA11	1.17	0.003	Annexin_A11	P:immune response
IPI00218946.2	HCN2	-1.24	0.003	Potassium/sodium_hyperpolarization-activated_cyclic_nucleotide-gated_channel_2	P:muscle contraction P:cell-cell signaling P:cation transport
IPI00304409.3	CARHSP1	1.20	0.003	Calcium-regulated_heat_stable_protein_1	P:intracellular signaling cascade
IPI00646062.3	KLHDC8A	1.17	0.003	Kelch_domain-containing_protein_8A	
IPI00030739.1	APOM	-1.14	0.003	Apolipoprotein_M	P:membrane lipid metabolic process P:generation of precursor metabolites and energy
IPI00152424.1	SLC2A12	1.16	0.004	solute_carrier_family_2_(facilitated_glucose_transporter),_member_12	
IPI00418802.1	LOC400968	-1.14	0.004	CDNA_FLJ45884_fis,_clone_OCBBF3021166	
IPI00030939.3	GNAS	-1.16	0.005	Alpha_subunit_of_GsGTP_binding_protein	P: G-protein coupled receptor protein signalin...
IPI00383161.1		1.28	0.005	Serine/threonine_protein_kinase_kkialre-like_1	
IPI00164387.1	KCNMA1	1.16	0.005	Isoform_5_of_Calcium-activated_potassium_channel_subunit_alpha_1	P:synaptic transmission P:potassium ion transport
IPI00744443.1		1.53	0.005	Similar_to_calcium_channel,_voltage-dependent,_alpha_2/delta_3_subunit	
IPI00008868.3	MAP1B	-1.23	0.005	Microtubule-associated_protein_1B	
IPI00395834.3		1.14	0.005	116_kDa_protein	
IPI00479962.3	MYO5B	-1.09	0.006	Myosin-5B	
IPI00640963.1	NAP1L6	-1.21	0.006	12_kDa_protein	
IPI00030255.1	PLOD3	1.28	0.006	Procollagen-lysine,2-oxoglutarate_5-dioxygenase_3_precursor	P:protein modification process
IPI00152422.1	FLJ42875	-1.16	0.006	Hypothetical_protein_DKFZp761G0122	
IPI00167908.3	IQUB	1.10	0.006	FLJ35834_protein	
IPI00386442.5	KIFAP3	1.14	0.006	Kinesin-associated_protein_3	P:signal transduction P:protein complex assembly P:microtubule-based process
IPI00329444.3	ACSM2B	-1.15	0.008	Hypothetical_protein_HYST1046	
IPI00015963.2	TRPA1	-1.24	0.008	Transient_receptor_potential_cation_channel_subfamily_A_member_1	P:ion transport
IPI00065931.2	AKAP13	1.10	0.008	A-kinase_anchor_protein_13_isoform_1	

IPI00185919.3	LARP1	-1.23	0.009	Isoform_1_of_La-related_protein_1	
IPI00012451.2	GNB4	1.09	0.009	Guanine_nucleotide-binding_protein_subunit_beta_4	
IPI00166009.2	FBXL11	-1.22	0.009	Isoform_1_of_JmjC_domain-containing_histone_demethylation_protein_1A	
IPI00442880.1	Q6ZP64	1.15	0.009	Hypothetical_protein_FLJ26451	
IPI00306604.5	ITGA5	-1.14	0.009	Integrin_alpha-5_precursor	P:cell adhesion
IPI00740057.2	ANKRD31	-1.19	0.009	PREDICTED:_similar_to_Ankyrin_repeat_domain-containing_protein_11	
IPI00216061.3	PRR5	-1.19	0.009	proline_rich_5_(renal)_isoform_2	
IPI00011919.1	FADD	-1.15	0.010	FADD_protein	P: positive regulation of I-kappaB kinase/NF-k... P: induction of apoptosis via death domain rec...
8923808	EPHA5	-1.21	0.010	erythropoietin_4_immediate_early_response_[Homo_sapiens]	
IPI00787100.1	LOC728328	-1.14	0.010	PREDICTED:_similar_to_procollagen,_type_I,_alpha_2	
18086504	CYP4B1	1.15	0.011	cytochrome_P450_[Homo_sapiens]	
IPI00017672.4		1.18	0.011	Hypothetical_protein_FLJ25678	
IPI00166002.3	PSD3	-1.14	0.011	PSD3_protein	
1684923		-1.41	0.011	This_CDS_feature_is_included_to_show_the_translation_of_the_corresponding_V_region,_partial_cds_Presently_translation_qualifiers_on_V_region_features_are_illegal	
13543559	NDUFA12 (DAP13)	1.14	0.011	NADH_dehydrogenase_(ubiquinone)_1_alpha_subcomplex,_12_[Homo_sapiens]	P:response to oxidative stress P:respiratory gaseous exchange
IPI00032358.3	POM121	-1.48	0.011	Nuclear_envelope_pore_membrane_protein_POM_121	
IPI00470901.3	PPP2R2A	-1.36	0.011	CDNA_FLJ41613_fis,_clone_CTONG3002412,_moderately_similar_to_Human_DOCK180_protein_mRNA.	
IPI00397393.2	KIAA1549	-1.38	0.011	PREDICTED:_similar_to_K06A9.1b_isoform_2	

34535254	DNAH17	1.09	0.013	unnamed_protein_product_[Homo_sapiens]	
IPI00413385.1	ITIH5L	-1.31	0.013	ITI-like_protein	
IPI00410324.1	LSM12	-1.31	0.013	LSM12_homolog	
IPI00019312.2	G7C	1.14	0.013	Protein_G7c_precursor	
IPI00302660.2	VPS37A	1.17	0.013	Vacuolar_protein_sorting_37_homolog_A	
IPI00142538.3	SETX	1.17	0.013	Isoform_1_of_Probable_helicase_senataxin	P:RNA processing P:double-strand break repair
IPI00030045.2	PROK2	1.17	0.013	Isoform_1_of_Prokineticin-2_precursor	P: spermatogenesis P: sensory perception of pain P: positive regulation of smooth muscle contra... P: inflammatory response P: G-protein coupled receptor protein signalin... P: elevation of cytosolic calcium ion concentr... P:chemotaxis P:cell proliferation P:anti-apoptosis P:angiogenesis P:activation of MAPK activity
8017376		-1.21	0.014	sphingosine_kinase_[Homo_sapiens]	P: sphingosine metabolic process P: sphingoid catabolic process P: positive regulation of smooth muscle contra... P: positive regulation of progression through ... P: positive regulation of fibroblast proliferation P: positive regulation of cell migration P: positive regulation of cell growth P: positive regulation of angiogenesis P: G-protein coupled receptor protein signalin... P:calcium-mediated signaling P:anti-apoptosis
IPI00103869.1	CTTNBP2	1.22	0.015	Cortactin-binding_protein_2	
IPI00060544.3	DDIT4L	-1.12	0.015	DNA-damage-inducible_transcript_4-like	
IPI00444179.2	LSDP5	1.27	0.015	PREDICTED:_similar_to_lipid_droplet_associated_protein	
IPI00787962.1	LRRC48	-1.18	0.015	PREDICTED:_similar_to_leucine_rich_repeat_containing_48	
IPI00328156.8	MAOB	-1.18	0.015	Amine_oxidase_[flavin-containing]_B	

IPI00003926.2	CLN8	-1.17	0.015	Protein_CLN8	P:protein catabolic process P:phospholipid metabolic process P:nervous system development P:negative regulation of proteolysis P:lipid transport P:cholesterol metabolic process P:ceramide biosynthetic process
IPI00064462.2	ABHD13	-1.17	0.015	hypothetical_protein_LOC84945	
IPI00152637.1	ADAMTS15	-1.25	0.016	ADAMTS-15_precursor	
IPI00152627.4	C11orf30	-1.17	0.016	Chromosome_11_open_reading_frame_30	
IPI00549189.3	THOP1	-1.17	0.016	Thimet_oligopeptidase	
IPI00745599.1	NBLA00301	-1.13	0.016	Similar_to_Heart-_and_neural_crest_derivatives-expressed_protein_2	
IPI00180515.2	NBEAL2	1.14	0.016	CDNA_FLJ35552_fis,_clone_SPLEN2004346	
IPI00020668.4	UTF1	-1.13	0.016	UTF1	P: positive regulation of transcription from R...
IPI00023287.4	NEDD4L	-1.13	0.016	Isoform_4_of_E3_ubiquitin-protein_ligase_NEDD4-like_protein	P:water homeostasis P:sodium ion transport P:response to metal ion P:regulation of protein catabolic process P:protein ubiquitination P:positive regulation of endocytosis P:excretion P:cellular sodium ion homeostasis
IPI00104698.5		1.16	0.017	31_kDa_protein	
IPI00384597.1	PHC2	-1.15	0.017	Polyhomeotic_2_homolog	
IPI00151888.4	DOCK9	-1.11	0.017	Isoform_1_of_Dedicator_of_cytokinesis_protein_9	
IPI00787064.1		1.16	0.018	PREDICTED:_similar_to_statin-like	
IPI00032056.3	C17orf85	-1.14	0.018	ELG_protein	
IPI00017800.2	ABCA3	-1.14	0.018	ATP-binding_cassette_sub-family_A_member_3	P:transport P:response to drug
IPI00033054.4	CTDSPL2	-1.21	0.019	CTD_(carboxy-terminal_domain,_RNA_polymerase_II,_polypeptide_A)_small_phosphatase_like_2	
IPI00398169.2	LOC165186	-1.13	0.019	CDNA_FLJ43756_fis,_clone_TESTI2045920	
IPI00465243.4	TMC4	-1.13	0.019	Transmembrane_channel-like_protein_4	
6912632		1.15	0.019	rearranged_L-myc_fusion_sequence_[Homo_sapiens]	
IPI00177655.2	LRRC42	-1.13	0.019	51_kDa_protein	

IPI00162549.2	SP110	-1.32	0.019	SP110_nuclear_body_protein_isoform_c	
IPI00009737.1	RRAGD	1.13	0.019	Isoform_1_of_Ras-related_GTP-binding_protein_D	
5911869	DKFZp434E066	1.25	0.019	hypothetical_protein_[Homo_sapiens]	
IPI00171407.4	STXBP5	1.15	0.019	Isoform_1_of_Syntaxin-binding_protein_5	
IPI00640568.1	ALDOA	1.23	0.019	15_kDa_protein	
IPI00020002.3	CNO	-1.20	0.020	Protein_cappuccino_homolog	
IPI00293592.2	C16orf71	1.11	0.020	hypothetical_protein_LOC146562	
IPI00000375.5	CAPS2	1.15	0.020	Isoform_2_of_Calcyphosin-2	
IPI00478959.2	SPATA13	1.11	0.020	OTTHUMP00000018127	
IPI00026546.1	PAFAH1B2	1.11	0.020	Platelet-activating_factor_acetylhydrolase_IB_subunit_beta	P:lipid metabolic process
IPI00386113.1	C20orf59	1.13	0.021	Hypothetical_protein_FLJ23412	
IPI00218013.6	SGOL2	-1.21	0.021	Isoform_2_of_Shugoshin-like_2	
IPI00073357.1	FBXO4	1.11	0.021	Isoform_1_of_F-box_only_protein_4	P:protein ubiquitination
IPI00027627.1	TNFAIP8	1.11	0.021	MDC-3.13_isoform_2	P:negative regulation of anti-apoptosis
IPI00745619.1	FLJ27502	-1.20	0.021	Similar_to_37LRP/p40	
IPI00289334.1	FLNB	-1.21	0.021	Isoform_1_of_Filamin-B	P:signal transduction P:cytoskeletal anchoring P:actin cytoskeleton organization and biogenesis
IPI00216856.3	ANKMY2	1.13	0.021	Ankyrin_repeat_and_MYND_domain-containing_protein_2	
IPI00419908.3	GPR179	-1.20	0.022	Probable_G-protein_coupled_receptor_179_precursor	
IPI00009146.3	TRAFD1	1.11	0.022	Fln29	
IPI00003905.1	SLC2A2	-1.19	0.023	Solute_carrier_family_2,_facilitated_glucose_transporter_member_2	P:glucose transport P:carbohydrate metabolic process
IPI00160622.1	CEP250	1.35	0.023	Isoform_1_of_Centrosome-associated_protein_CEP250	P:regulation of centriole-centriole cohesion P:mitotic cell cycle
IPI00007425.1	DSC1	-1.22	0.023	desmocollin_1_isoform_Dsc1b_preproprotein	P:homophilic cell adhesion
IPI00030941.5	TSPAN3	-1.12	0.023	Tetraspanin-3	
7020225		1.12	0.023	unnamed_protein_product_[Homo_sapiens]	

IPI00289116.2	GPAM	-1.21	0.023	Hypothetical_protein_FLJ23960	P:phospholipid biosynthetic process
IPI00026663.2	ALDH1A3	-1.20	0.023	Aldehyde_dehydrogenase_1A3	P:lipid metabolic process P:alcohol metabolic process
IPI00006640.3	SERPINI2	1.11	0.023	Serpin_I2_precursor	P:cell motility
IPI00642716.3	MYH7B	-1.20	0.023	OTTHUMP00000030720	
IPI00217690.1	MGC42105	-1.10	0.023	Serine/threonine-protein_kinase_NIM1	P:protein amino acid phosphorylation
IPI00007626.1	KCNJ4	-1.20	0.023	Inward_rectifier_potassium_channel_4	P:potassium ion transport
IPI00746049.1	ANKRD30B	-1.14	0.023	Similar_to_Breast_cancer_antigen_NY-BR-1.1	
52693921	SHC1	1.22	0.023	SHC_(Src_homology_2_domain_containing)_transforming_protein_1_isoform_p66Shc_[Homo_sapiens]	P: regulation of epidermal growth factor recep... P:positive regulation of mitosis P:positive regulation of cell proliferation P:activation of MAPK activity
IPI00021326.4	SHC1	1.22	0.023	SHC_(Src_homology_2_domain_containing)_transforming_protein_1	P:intracellular signaling cascade
IPI00217696.3	PPIL6	-1.30	0.023	Peptidyl-prolyl_cis-trans_isomerase	
IPI00383200.1		1.13	0.023	HSPC271_(Fragment)	
IPI00005559.1	NOD2	1.10	0.023	Isoform_1_of_Caspase_recruitment_domain-containing_protein_15	P: protein oligomerization P: positive regulation of interleukin-1 beta s... P: positive regulation of I-kappaB kinase/NF-k... P:detection of muramyl dipeptide P:detection of bacterium P:defense response to bacterium P:activation of NF-kappaB transcription factor
IPI00418236.1	FIGLA	1.10	0.023	Factor_in_the_germline_alpha	P:oocyte development
IPI00103595.2	CEP350	1.11	0.023	Centrosome-associated_protein_350	
IPI00739552.2		1.18	0.023	Pseudogene_candidate	
IPI00006714.2	PTPN13	-1.19	0.024	Isoform_3_of_Tyrosine-protein_phosphatase_non-receptor_type_13	P:protein amino acid dephosphorylation
IPI00162733.2	KCNC2	1.10	0.024	Kv3.2d_voltage-gated_potassium_channel	
8926561		-1.18	0.024	AF274840_1_T_cell_receptor_beta_chain_[Homo_sapiens]	
IPI00017974.1	9-Sep	-1.20	0.024	CDNA_FLJ12190_fis,_clone_MAMMA1000842	
IPI00012504.2	ZNF44	-1.20	0.024	zinc_finger_protein_44	

IPI00472977.2	USP37	1.09	0.024	Ubiquitin_carboxyl-terminal_hydrolase_37	
IPI00024074.1	CCDC69	-1.28	0.025	CDNA_FLJ13705_fis,_clone_PLACE2000302	
IPI00428657.1	GIGYF1	1.10	0.025	PERQ_amino_acid_rich,_with_GYF_domain_1	
IPI00302010.5	FLJ20309	-1.13	0.026	Hypothetical_protein_FLJ20309	
IPI00374484.4	LOC731911	1.20	0.027	PREDICTED:_hypothetical_protein	
IPI00289877.7	TOX3	-1.14	0.027	PREDICTED:_similar_to_trinucleotide_repeat_containing_9_isoform_4	
IPI00786918.1	FAM82B	-1.17	0.027	hypothetical_protein_LOC51115	
IPI00175977.3	TDRD5	1.09	0.027	CDNA_FLJ45446_fis,_clone_BRSSN2015497	
IPI00384202.3	JMJD1C	1.09	0.027	Isoform_1_of_Probable_JmjC_domain-containing_histone_demethylation_protein_2C	P:regulation of transcription, DNA-dependent
IPI00217481.2	GPR126	1.48	0.027	Developmentally_regulated_G-protein-coupled_receptor_beta_1	P: G-protein coupled receptor protein signalin...
IPI00382872.1	DUX4	1.21	0.027	Facioscapulohumeral_muscular_dystrophy	P:regulation of transcription, DNA-dependent
IPI00099871.1	MRPL40	-1.19	0.027	39S_ribosomal_protein_L40,_mitochondrial_precursor	P:anatomical structure morphogenesis
IPI00025100.1	BCKDHA	-1.13	0.027	2-oxoisovalerate_dehydrogenase_alpha_subunit,_mitochondrial_precursor	P:branched chain family amino acid catabolic ...
IPI00304742.4	STK10	-1.13	0.027	113_kDa_protein	P:protein amino acid phosphorylation
IPI00744046.1		-1.13	0.027	Similar_to_Glutamate_receptor_delta-1_subunit_precursor	
IPI00299095.2	SNX2	-1.10	0.027	Sorting_nexin-2	P:endocytosis
IPI00646354.1	PTAR1	1.15	0.027	OTTHUMP00000021436	P:protein amino acid prenylation
IPI00384443.4	BRD9	-1.21	0.028	Sarcoma_antigen_NY-SAR-29	
IPI00060310.4	PLD4	1.21	0.028	phospholipase_D_family,_member_4	
IPI00163782.2		1.18	0.028	Isoform_2_of_Far_upstream_element-binding_protein_1	
IPI00005210.1	CETN3	1.10	0.028	Centrin-3	P:centrosome cycle
IPI00017596.2	MAPRE1	-1.13	0.028	Microtubule-associated_protein_RP/EB_family_member_1	P: regulation of progression through cell cycle P: negative regulation of microtubule polymeri... P:cell proliferation

20151073		-1.23	0.029	A_Chain_A,_Human_Tissue_Transglutaminase_In_Gdp_Bound_Form	
34527259		1.19	0.029	unnamed_protein_product_[Homo_sapiens]	
10432782	TES	-1.19	0.029	unnamed_protein_product_[Homo_sapiens]	
IPI00033036.1	METAP2	1.08	0.029	Methionine_aminopeptidase_2	P:protein processing P:peptidyl-methionine modification P:N-terminal protein amino acid modification
IPI00394857.1	EFCAB5	-1.33	0.029	CDNA_FLJ46247_fis,_clone_TESTI4021129	
IPI00023757.2	RPGR	1.09	0.030	Isoform_1_of_X-linked_retinitis_pigmentosa_GTPase_regulator	P:visual perception P:intracellular protein transport
IPI00221224.4	ANPEP	1.13	0.031	Aminopeptidase_N	
IPI00034099.3	RBM35B	1.07	0.031	Hypothetical_protein_FLJ21918	
IPI00006093.2	FAM38A	-1.16	0.031	Protein_FAM38A	
IPI00604745.2	MAPK13	1.14	0.031	MAPK13_protein_variant	P:protein amino acid phosphorylation
IPI00001348.5	SFI1	1.24	0.031	spindle_assembly_associated_Sfi1_homolog_isoform_a	
IPI00255653.4	ATP11A	-1.13	0.033	Probable_phospholipid-transporting_ATPase_IH	
IPI00299024.8	BASP1	1.10	0.033	Brain_acid_soluble_protein_1	
IPI00103586.5	CPA5	-1.13	0.033	Isoform_1_of_Carboxypeptidase_A5_precursor	
IPI00176532.1	JPH2	1.19	0.033	Isoform_1_of_Junctophilin-2	
IPI00220956.2	MTMR1	-1.19	0.033	Isoform_1A_of_Myotubularin-related_protein_1	
IPI00303401.4	C1orf75	-1.10	0.034	FLJ10874_protein	
IPI00015697.3	C13orF14	-1.19	0.034	OTTHUMP00000018495	
48146179	IFIT5	-1.08	0.035	IFIT5_[Homo_sapiens]	
IPI00375803.3	GON4L	-1.19	0.036	Isoform_1_of_GON-4-like_protein	
IPI00012322.1	PRKG2	-1.11	0.036	cGMP-dependent_protein_kinase_2	P:signal transduction P:regulation of progression through cell cycle P:protein amino acid phosphorylation
IPI00032063.5	LRP1B	-1.21	0.036	CDNA_FLJ30101_fis,_clone_BNGH41000118,_highly_similar_to_Homo_sapiens_low_density_lipoprotein_receptor_related_protein-deleted_in_tumor_(Fragment)	

IPI00329591.3	RMND1	-1.18	0.037	Isoform_1_of_Protein_C6orf96	
IPI00215827.1	AMPD2	-1.12	0.038	Isoform_Ex1B-3_of_AMP_deaminase_2	P:purine nucleotide metabolic process
IPI00017469.1	SPR	-1.12	0.038	Sepiapterin_reductase	P:tetrahydrobiopterin biosynthetic process P:nitric oxide biosynthetic process P:electron transport
IPI00167903.5	ZNF555	-1.19	0.039	Isoform_2_of_Zinc_finger_protein_555	
IPI00010604.4	PLCE1	1.13	0.040	Phospholipase_C,_epsilon_1	
IPI00154489.1	LOC146325	-1.19	0.040	Hypothetical_protein_RJD1	
IPI00640218.1	ZNF709	-1.19	0.040	Zinc_finger_protein_709	
IPI00241409.8	FAM21B	-1.17	0.040	PREDICTED:_similar_to_CG16742-PA,_isoform_A_isoform_1	
IPI00398791.1	LOC375127	-1.21	0.040	Hypothetical_protein_FLJ26056	
IPI00009329.1	UTRN	1.13	0.040	Utrophin	P:muscle development P:muscle contraction
IPI00177519.2	RBM24/RBM38/RNPC1	1.22	0.040	24_kDa_protein	
22760635	SHC1	-1.12	0.040	unnamed_protein_product_[Homo_sapiens]	
IPI00011932.7	HSPA12A	1.12	0.040	PREDICTED:_similar_to_Heat_shock_70_kDa_protein_12A_isoform_1	
IPI00008998.1	PTPLAD1	-1.15	0.040	HSPC121	
IPI00148061.3	LDHAL6A	1.12	0.040	L-lactate_dehydrogenase_A-like_6A	
IPI00240059.3	TMCC3	-1.15	0.041	Transmembrane_and_coiled-coil_domains_protein_3	
IPI00329637.2	C1orf26	-1.14	0.041	Uncharacterized_protein_C1orf26	
IPI00167788.1	C16orf81	-1.12	0.042	CDNA_FLJ36701_fis,_clone_UTERU2009147	
IPI00168501.3	ZC3H14	1.11	0.042	nuclear_protein_UKp68_isoform_3	
IPI00025276.1	TNXB	1.13	0.043	Isoform_XB_of_Tenascin-X_precursor	P:elastic fiber assembly P:collagen metabolic process P:cell adhesion P:actin cytoskeleton organization and biogenesis
IPI00015826.1	ABCB10	-1.19	0.043	ATP-binding_cassette_sub-family_B_member_10,_mitochondrial_precursor	P:transport

IPI00419849.3	C19orf2	-1.11	0.043	RNA_polymerase_II_subunit_5-mediating_protein	P: response to virus P: regulation of transcription from RNA polyme...
IPI00738733.2	LOC643491	1.19	0.045	PREDICTED:_similar_to_Golgin_subfamily_A_member_2	
IPI00291755.5	NUP210	-1.19	0.045	206_kDa_protein	
627616		1.12	0.047	PH0229_T-cell_receptor_Vb_CDR3,_carrier_Vb_17.sbt_-_human_(fragment)	
IPI00007993.4	HIC1	-1.11	0.048	Isoform_1_of_Hypermethylated_in_cancer_1_protein	P:regulation of transcription, DNA-dependent P:multicellular organismal development
IPI00021034.1	COL4A1	-1.20	0.048	Collagen_alpha-1(IV)_chain_precursor	
29646899		-1.16	0.048	T_cell_receptor_beta_chain_[Homo_sapiens]	
IPI00738065.1	LOC652627	-1.11	0.048	PREDICTED:_similar_to_dead_end_homolog_1	
IPI00410214.1	BPNT1	-1.16	0.050	Isoform_1_of_3'(2'),5'-bisphosphate_nucleotidase_1	P: nucleobase, nucleoside, nucleotide and nucl... P:nervous system development
IPI00513975.2	PRRT1	1.16	0.050	Chromosome_6_open_reading_frame_31	

- Q-Values for both protein identification and quantification are the minimal q-values from all samples.
- Fold change= Mean_T / Mean _C when Mean_T ≥ Mean_C (up-regulation)
  Negative values mean down-regulation in the cancer plasma sample.
- All protein with significant changes were annotated and categorized based on their biological function with Gene Ontology.

Appendix 2: Luminal B (P4) enriched proteins with significant changes

Protein_ID	Annotation	q_H_P4	FC_H_P4
IPI00477597.1	Isoform_1_of_Haptoglobin-related_protein_precursor	7.52E-05	1.96
IPI00431645.1	HP_protein	7.52E-05	2.12
IPI00478493.3	HP_protein	7.52E-05	2.04
IPI00641737.1	Haptoglobin_precursor	7.52E-05	2.07
306882	haptoglobin_precursor	7.52E-05	2.06
IPI00465006.1	Hypothetical_protein_DKFZp686D19113	0.0004134	2.03
IPI00021727.1	C4b-binding_protein_alpha_chain_precursor	0.0023582	1.31
IPI00025862.1	C4b-binding_protein_beta_chain_precursor	0.0037023	1.26
229386	720005A_protein,alpha1_acid_glyco	0.0037023	2.05
IPI00294004.1	Vitamin_K-dependent_protein_S_precursor	0.0038465	1.27
IPI00032056.3	ELG_protein	0.0059168	-1.69
IPI00431656.4	Isoform_2_of_Alpha-1-antichymotrypsin_precursor	0.0064549	1.33
IPI00022429.3	Alpha-1-acid_glycoprotein_1_precursor	0.0064549	1.77
48425723	E_Chain_E,_Structure_Of_Human_Transferrin_Receptor-Transferrin_Complex	0.0064549	-1.56
IPI00025204.1	CD5_antigen-like_precursor	0.006576	-1.26
IPI00550991.2	Isoform_1_of_Alpha-1-antichymotrypsin_precursor	0.006576	1.28
225769	1313184C_chymotrypsin_inhibitor	0.006576	1.29
IPI00021364.1	Properdin_precursor	0.006576	-1.2
IPI00022392.1	Complement_C1q_subcomponent_subunit_A_precursor	0.0089154	-1.21
126608	LYSOZYME_Spiked_Standard_(HEN)	0.0089154	-1.17
IPI00011264.1	Complement_factor_H-related_protein_1_precursor	0.0089154	1.27
IPI00022463.1	Serotransferrin_precursor	0.0089154	-1.4
7245523	A_Chain_A,_Human_Serum_Transferrin	0.0133428	-1.4
IPI00020091.1	Alpha-1-acid_glycoprotein_2_precursor	0.0158175	1.62
IPI00216438.3	Solute_carrier_family_12_member_3	0.0163767	1.47
IPI00022391.1	Serum_amyloid_P-component_precursor	0.0168972	1.3
IPI00027507.1	Complement_factor_H-related_protein_3_precursor	0.0172369	1.49
IPI00785067.1	Hypothetical_protein	0.0173138	-1.42
IPI00642716.3	OTTHUMP00000030720	0.0173138	-1.78
IPI00015697.3	OTTHUMP00000018495	0.0173138	-1.81
IPI00218013.6	Isoform_2_of_Shugoshin-like_2	0.0173138	-1.83
IPI00164672.5	mRNA_decapping_enzyme_1A	0.0173138	-1.99
IPI00301072.3	Small_conductance_calcium-activated_potassium_channel_protein_2	0.0173138	-1.41
IPI00441894.1	Hypothetical_protein_FLJ24000	0.0173138	1.81
IPI00032328.1	Isoform_HMW_of_Kininogen-1_precursor	0.0236039	-1.14
IPI00027462.1	Protein_S100-A9	0.0250229	1.37
21707947	Leucine-rich_alpha-2-glycoprotein_1_[Homo_sapiens]	0.0279059	1.35
IPI00007047.1	Protein_S100-A8	0.0279853	1.43
IPI00167093.4	Complement_factor_H-related_1	0.0283429	1.32
IPI00021891.5	Isoform_Gamma-B_of_Fibrinogen_gamma_chain_precursor	0.0289841	1.33
IPI00654888.2	Kallikrein_B,_plasma_(Fletcher_factor)_1	0.0298312	-1.13
IPI00006987.1	ATP-dependent_RNA_helicase_DDX24	0.0310875	1.47

1575607	FUSE_binding_protein_2_[Homo_sapiens]	0.0310875	-1.7
IPI00001611.1	Isoform_1_of_Insulin-like_growth_factor_II_precursor	0.0310875	-1.23
IPI00030013.1	Homeobox_protein_SIX3	0.0310875	-1.65
IPI00018583.3	Hyaluronan_binding_protein_4	0.0310875	1.23
IPI00742996.1	IMP_dehydrogenase/GMP_reductase_family_protein	0.0311546	-1.65
IPI00303963.1	Complement_C2_precursor_(Fragment)	0.0323115	1.11
IPI00027410.1	Platelet_glycoprotein_V_precursor	0.0357933	-1.42
IPI00445774.1	CDNA_FLJ43404_fis,_clone_OCBBF2017516	0.0382753	-1.5
IPI00020996.3	Insulin-like_growth_factor-binding_protein_complex_acid_labile_chain_precursor	0.0390292	-1.16
IPI00298497.3	Fibrinogen_beta_chain_precursor	0.0390292	1.35
223002	0401173A_fibrin_beta	0.0390292	1.36
IPI00183706.3	KIF19_protein_(Fragment)	0.04351	1.36
IPI00022395.1	Complement_component_C9_precursor	0.0440995	1.17
IPI00398021.1	27_kDa_protein	0.0452365	1.4
IPI00215894.1	Isoform_LMW_of_Kininogen-1_precursor	0.0453133	-1.13
IPI00292218.3	Hepatocyte_growth_factor-like_protein_precursor	0.0453133	1.23
IPI00004656.1	Beta-2-microglobulin_precursor	0.0465936	-1.14
IPI00018305.3	Insulin-like_growth_factor-binding_protein_3_precursor	0.0480621	-1.24
IPI00787962.1	PREDICTED:_similar_to_leucine_rich_repeat_containing_48	0.0480621	-1.34
IPI00394851.1	hypothetical_protein_LOC346689	0.0480621	-1.27
IPI00788062.1	PREDICTED:_similar_to_protein_immuno-reactive_with_anti-PTH_polyclonal_antibodies	0.0484759	1.34
IPI00168529.2	hypothetical_protein_LOC126859_isoform_2	0.0485396	-1.46

Appendix 3: Proteins with significant changes enriched in Basal type (P1)

Protein_ID	Annotation	q_H_P1	FC_H_P1
IPI00465006.1	Hypothetical_protein_DKFZp686D19113	0.03681419	1.43
IPI00020091.1	Alpha-1-acid_glycoprotein_2_precursor	0.037612552	1.41
IPI00431645.1	HP_protein	0.033220081	1.41
IPI00027507.1	Complement_factor_H-related_protein_3_precursor	0.025613945	1.41
IPI00478493.3	HP_protein	0.033220081	1.4
306882	haptoglobin_precursor	0.033220081	1.39
IPI00641737.1	Haptoglobin_precursor	0.033220081	1.39
IPI00477597.1	Isoform_1_of_Haptoglobin-related_protein_precursor	0.033220081	1.37
IPI00216438.3	Solute_carrier_family_12_member_3	0.03681419	1.31
IPI00025862.1	C4b-binding_protein_beta_chain_precursor	0.001962815	1.22
IPI00019399.1	Serum_amyloid_A-4_protein_precursor	0.033220081	1.19
IPI00021727.1	C4b-binding_protein_alpha_chain_precursor	0.025613945	1.19
37947	unnamed_protein_product_[Homo_sapiens]	0.033220081	1.18
IPI00023014.1	von_Willebrand_factor_precursor	0.033220081	1.18
IPI00026314.1	Isoform_1_of_Gelsolin_precursor	0.047272606	-1.15
IPI00299040.1	Polycystin-2	0.037802977	-1.16
IPI00001611.1	Isoform_1_of_Insulin-like_growth_factor_II_precursor	0.03681419	-1.17
IPI00025204.1	CD5_antigen-like_precursor	0.025613945	-1.2
IPI00022431.1	Alpha-2-HS-glycoprotein_precursor	0.04335994	-1.22
IPI00218539.3	Isoform_B_of_Collagen_alpha-1(XI)_chain_precursor	0.042958593	-1.26
IPI00301072.3	Small_conductance_calcium-activated_potassium_channel_protein_2	0.03681419	-1.28
IPI00785067.1	Hypothetical_protein	0.03681419	-1.3
IPI00032056.3	ELG_protein	0.04736649	-1.32
IPI00019451.3	MRG-binding_protein	0.04736649	-1.32
8885790	AF146692_1_filamin_2_[Homo_sapiens]	0.03681419	-1.35
IPI00328762.4	ATP_binding_cassette,_sub-family_A_(ABC1),_member_13	0.040977469	-1.35
IPI00465430.5	70_kDa_protein	0.04736649	-1.36
IPI00010700.2	Isoform_1_of_Large_proline-rich_protein_BAT2	0.037802977	-1.36
1017427	elastic_titin_[Homo_sapiens]	0.04736649	-1.37
IPI00445774.1	CDNA_FLJ43404_fis,_clone_OCBBF2017516	0.03681419	-1.37
IPI00742996.1	IMP_dehydrogenase/GMP_reductase_family_protein	0.03681419	-1.47
IPI00218013.6	Isoform_2_of_Shugoshin-like_2	0.03681419	-1.5
IPI00642716.3	OTTHUMP00000030720	0.03681419	-1.51
IPI00015697.3	OTTHUMP00000018495	0.03681419	-1.58

Appendix 4: Significance differences among Luminal type B and Basal

Protein_ID	Annotation	q_H_p4	FC_p1_p4
IPI00431645.1	HP_protein	7.52E-05	1.50
IPI00641737.1	Haptoglobin_precursor	7.52E-05	1.49
306882	haptoglobin_precursor	7.52E-05	1.48
IPI00478493.3	HP_protein	7.52E-05	1.46
IPI00465006.1	Hypothetical_protein_DKFZp686D19113	0.000413398	1.42
IPI00477597.1	Isoform_1_of_Haptoglobin-related_protein_precursor	7.52E-05	1.43
IPI00006987.1	ATP-dependent_RNA_helicase_DDX24	0.031087485	1.36
IPI00007047.1	Protein_S100-A8	0.027985303	1.42
IPI00398021.1	27_kDa_protein	0.045236465	1.41
IPI00027462.1	Protein_S100-A9	0.025022886	1.33
IPI00788062.1	PREDICTED:_similar_to_protein_immuno-reactive_with_anti-PTH_polyclonal_antibodies	0.048475938	1.31
IPI00167093.4	Complement_factor_H-related_1	0.028342855	1.29
IPI00011264.1	Complement_factor_H-related_protein_1_precursor	0.00891538	1.22
IPI00292218.3	Hepatocyte_growth_factor-like_protein_precursor	0.045313251	1.26
IPI00215894.1	Isoform_LMW_of_Kininogen-1_precursor	0.045313251	-1.12
IPI00004656.1	Beta-2-microglobulin_precursor	0.046593554	-1.15
IPI00032328.1	Isoform_HMW_of_Kininogen-1_precursor	0.0236039	-1.13
IPI00021364.1	Properdin_precursor	0.006576016	-1.14
IPI00394851.1	hypothetical_protein_LOC346689	0.048062145	-1.27
IPI00787962.1	PREDICTED:_similar_to_leucine_rich_repeat_containing_48	0.048062145	-1.30
IPI00022463.1	Serotransferrin_precursor	0.00891538	-1.26
48425723	E_Chain_E,_Structure_Of_Human_Transferrin_Receptor-Transferrin_Complex	0.00645493	-1.32
IPI00030013.1	Homeobox_protein_SIX3	0.031087485	-1.69

References:

1. Jemal A, et al. Cancer statistics. CA Cancer J Clin. 2006, 56(2):106-30.

2. Retrieved May 2008 from www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the _key_statistics.

3. Jemal A, Tiwari RC, Murray T, et al. Cancer statistics. CA Cancer J Clin. 2004, 54(8):18-29.

4. Wulfkuhle JD, et al. Proteomic applications for the early detection of cancer. Nature reviews/cancer. April 2003, 3(4):267-75.

5. Parkin DM, Bray F, Ferlay J, Pisani P. Estimating the world cancer burden. Globocan Int J Cancer. 2001, 94:153-56.

6. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics. CA Cancer J Clin. 2005, 55:74-108.

7. Retrieved May 2008 from http://www.cancer.org/docroot/CRI/content/CRI_2_4_3X_Can_breast_cancer_be_found_early_5.asp. Can breast cancer be found early?

8. Antman K, Shea S. Screening mammography under age 50. JAMA. 1999, 281:1470-1472.

9. Hondermarck H, et al. Breast Cancer when proteomics challenges biological complexity. Molecular & Cellular Proteomics. 2003, 2:281-291.

10. Retrieved December 2007 from http://www.enotalone.com/article/8598.html. Limitations of Mammography, When to Get a Mammogram?

11. Sorlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implication. Proc Natl Acad Sci USA. 2001, 98(19):10869-74.

12. Srinivas PR, Kramer BS, Srivastava S. Trends in biomarker research for cancer detection. Lancet Oncol. 2001, 2:698-704.

13. Pothur R, et al. Proteomics for cancer biomarker discovery. Clinic Chemistry. 2002, 48-8: 1160-69.

14. Polanski M, Anderson N.L. A List of Candidate Cancer Biomarkers. Biomarker Insight. 2006, 5:21-48.

15. Mor G, Visintin I, Lai Y. Serum protein markers for early detection of ovarian cancer. PNAS. 2005, Vol. 102 No.21:7677-82.

16. Pothur R, et al. Proteomics for cancer biomarker discovery. Clinic Chemistry. 2002, 48-8: 1160-69.

17. O'Farrell PH et al. High resolution two-dimensional electrophoresis of proteins. J. Biochem 1975, 250:4007-4021.

18. Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Natl Biotech. 1999, 17:994-999.

19. Wei Y, Sharon S. Mass spectrometry-based quantitative proteomic profiling. Briefings in functional genomics and proteomics. May 2005, 4(1):27-38.

20. Higgs R, Knierman M. Comprehensive Label-Free Method for the Relative Quantification of Proteins from Biological Samples. Journal of Proteome Research. 2005, 4:1442-50.

21. Fitzpatrick D, You JS. Searching for potential biomarkers of cisplatin resistance in human ovarian cancer using a label-free LC/MS-based protein quantification method. Proteomics Clin. April 2007, (1):246-63.
22. Visintin G, Lai Y. Serum protein markers for early detection of ovarian cancer. PNAS. May 24 2005, 102 (21):7677-82.
23. Hinerfeld D, et al. Serum/Plasma Depletion with Chicken Immunoglobulen Y Antibodies for Proteomic Analysis from Multiple Mammalian Species. Journal of biomolecular techniques. 2004, 15(3):3721-3732.
24. Petricoin E, Liotta L. The vision for a new diagnostic paradigm. ClinChem. 2003, 49:1276-78.
25. Huang L. Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis. Proteomics 2005, 5:3314-28.
26. Retrieved April 2008 from http://www.genwaybio.com/gw_file.php?fid=2026
27. Hale JE, Butler JP, Gelfanova V, You J S, Knierman M D. A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. Anal. Biochem. 2004, 333:174-81.
28. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003, 19:185-93.
29. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. Molecular & Cellular Proteomics 2004, 3:531-33.

30. Retrieved May, 2008 from http://proteomics.cancer.gov/programs/CPTAC.

31. Retrieved September, 2007 from http://www.pir.uniprot.org.

32. Retrieved September, 2007 from http://www.geneontology.org.

33. Bradford M, et al. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding anal. Biochem 1976, 72:248-54.

34. Onitilo AA, Engel J. Breast cancer subtypes based on ER/PR and HER2 expression: A population-based experience. J Clin Oncol. 2008, 26: 325-332

35. Sorlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA 2001, 98(19):2639-48.

36. Retrieved September, 2008 from http://www.pwc.com/techforecast/pdfs/pharmaco-wb-x.pdf. The emerging pharmacogenomics revolution. Personalized Medicine.

37. Wagner K, Hemminki K. The insulin-like growth factor-1 pathway mediator genes: SHC1 Met 300Val shows a protective effect in breast cancer. Carcinogenesis. 2004, 25: 2473-2478

38. Schmid K, et al. Plasma proteins, structure, function and genetic control. Academic Press.1975, 1:183-288.

39. Irmak S, Tilki D. Stage-dependent increase of orosomucoid and zinc-alpha-glycoprotein in urinary bladder cancer. Proteomics 2005, 5:4296-304.

40. Surmacz E, et al. Function of the IGF-1 receptor in breast cancer. J. Mammary Gland Biol. Neoplasina. 2000, 5:95-105.

41. Ravichandran K, et al. Signaling via SHC family adapter proteins. Oncogene. 2001, 20:6322-30.

42. Stevenson, LE, Frackelton, ARJr. Constitutively tyrosine phosphorylated p52 Shc in breast cancer cells: Correlatin with ErbB2 and p66 Shc expression. Breast Cancer. 1998, 49:119-28.

43. Sivaraman V S, Wang H. Hyperexpression of mitogen-activated protein kinase in human breast cancer. J. Clin. Invest. 1997, 99:1478-83.

44. Saito S, Murayama Y. Haptoglobin-beta chain defined by monoclonal antibody RM2 as a novel serum marker for prostate cancer. Intl. J. Cancer. 2008 Aug 1, 123(3):633-40.

45. Ruosahti E, et al. The Walte Herbert Lecture. Control of cell motility and tumour invasion by extracellalar matrix interactions. Br. J. Cancer. 1992, 66:239-42.

46. Duband J L, et al. The instructive role of fibronectins in cell migrations during embryonic development. Ann. New York Acad. Sci. 1990, 588:273-80.

47. Balbona K, et al, Tran H. Fibulin binds to itself and to the carboxylterminal heparin-binding region of fibronectin. J. Biol. Chem. 1992, 267:20120-25.

48. Schwarzbauer JE, et al. Identification of the fibronectin sequences required for assembly of fibrillar matrix. J. Cell Biol. 1991, 113:1463-73.

49. Aquirre KM, et al. Fibronectin self-association is mediated by complementary sites within the amino-terminal one-third of the molecule. J. Biol. Chem. 1994, 269(45):27863-68.

50. Yi M, et al. Antiangiogenic proteins require plasma fibronectin or vitronectin for in vivo activity. Proc. Natl. Acad. Sci. USA. 2003, 100:11435-38.

51. Pang AL, et al. U 94 alters FN1 and ANGPTL4 gene expression and inhibits tumorigenesis of prostate cancer cell line PC3. Cancer cell international. 2005, 5:19 1475-2867.

52. Ye B, Cramer D W. Haptoglobin-α subunit as potential serum biomarker in ovarian cancer**.** Clinical cancer research 2003, 9:2904-11.

53. Goncalves A, Esterni B. Prostoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. Oncogene 2006, 25:981-89.

54. Sanders ME, Dias EC. Differentiating proteomic biomarkers in breast cancer by laser capture microdissection and MALDI MS. J. Proteome Res. 2008, *7* (4): 1500–1507.

55. Ehlermann P, Eggers K. Increased proinflammatory endothelial response to S100A8/A9 after preactivation through advanced glycation end products. Cardiovasc Diabetol 2006, 5:1475-2840.

56. Tolson JP, Flad T. Differential detection of S100A8 in transitional cell carcinoma of the bladder by pair wise tissue proteomic and immunohistochemical analysis. Proteomics 2003, 6(2): 697 – 708.

CURRICULUM VITAE

Guihong Qi

Education

Master of Science, Biotechnology (2008)
Indiana University, Indianapolis, IN
Thesis: "Biomarker Discovery in Early Stage Breast Cancer Using Proteomics Technologies"
Graduate Certificate in Biotechnology (2007)
Indiana University, Indianapolis, IN
Bachelor of Science, Chemistry (1988)
Harbin Normal University, Harbin, Heilongjiang, P.R.China

Professional Experience

Senior Research Associate, Monarch LifeSciences, LLC, Indianapolis, IN, 2004-present

- Worked on two-dimensional gel electrophoresis (2DE) based on proteomics projects, including sample preparation, protein assay, large format gradient and linear gels casting, iso-electric focusing, gel spots cutting and image analysis by using PDQuest software.
- Prepped serum, cell lysate, tissue, etc. biological samples for protein identification and quantification, including protein extraction, precipitation, dialysis, high abundant protein removal using affinity technique, in-solution and in-gel enzymatic digestion.
- Set up and maintained instruments (MALDI-TOF from Micromass, Q-TOF from Micromass, MALDI-TOF/TOF from Applied-Biosystems, LTQ from Thermofinnigan, spots cutter and Flour-S multi-imager from Bio-Rad) for proteins and peptides analysis including buffer preparation, calibration, tuning, cleaning and running.
- Performed database search for analyzed proteins using ProFound, Mascot and SEQUEST.
- Trained internal personnel and our clients.
- Interpreted results and documenting work with our scientist and customers.

Research Technician, Proteomics core facility, Department of Biochemistry and Molecular Biology, Indiana University, Indianapolis, IN, 2003-2004

- Prepared and run 2DE samples including extraction, desalting, protein assay, iso-electric focusing and SDS page.
- Cast large format gradient and linear gels.
- Scanned and analyzed 2DE images using PDQuest software.

- Performed manual and robotic protein digestion.
- Analyzed protein and peptide using MALDI-TOF.
- Performed protein peptide mass fingerprint.
- Generated data for our customers.

Research Technician, Department of Biochemistry and Molecular Biology, Indiana University, Indianapolis, IN, 2001-2003

- Synthesized peptide manually and by using ABI 431, 433-peptide synthesizer.
- Purified peptide with preparative HPLC.
- Characterized peptide by using analytical HPLC, TLC and Mass spectrometry.
- Independently performed peptide synthesis procedure.
- Recorded and monitored laboratory activities.

Research Experience

- Applied LC-MS/MS label free quantification technique in biomarker discovery and validation for pharmaceutical companies and academic clients.
- Analyzed biological pathways and find the proteins network connections from the protein dataset using Pathway Studio™.

Conferences Attended

NCI Annual Meeting on Clinical Proteomic Technologies for Cancer, 2008

- Poster session: "Quantitative Proteomic Analysis of Human Plasma Samples from Breast Cancer Patients Using an LC/MS-based Label-free Protein Quantification Platform" **Guihong Qi**, Jinsam You, Jong-Won Kim, Kerry Bemis and Mu Wang

Biochemistry and Molecular Biology Research Day, Indian University, Indianapolis, IN, 2008

- Poster session: "Biomarker Discovery in Early Stage Breast Cancer Using Proteomics Technologies" **Guihong Qi**, Jinsam You, Kerry Bemis and Mu Wang

The Indiana University School of Medicine biannual Dean's Grand Rounds and Scientific Session, 2008

- Poster session: "Biomarker Discovery in Early Stage Breast Cancer Using Proteomics Technologies" **Guihong Qi**, Jinsam You, Kerry Bemis and Mu Wang

Publications

Dawn P. G. Brown, **Guihong Qi**, Frank A. Witzmann, George W. Sledge Jr. , Mu Wang, "A comparative proteomic study to characterize the vinblastine resistance in human ovarian cancer cells", Journal of Proteomics, 2007. 1(1): p.18-31