Classification and Biomarker Discovery in Cancer Studies Using a Differential Protein Mapping Method Combined with Mass Spectrometry

by

Yanfei Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2008

Doctoral Committee:

Professor David M. Lubman, Chair
Professor Zhan Chen
Professor Kristina I. Hakansson
Professor Robert Zand

To my family.

Acknowledgements

I would like to thank my advisor Prof. David M. Lubman, for his encouragement and assistance through the course of my Ph.D study. Dr. Lubman's help over the last five years has been indispensable, both for providing facilities and financial support as well as his input on research. It was a precious experience working under his guidance. Sincerest appreciation also goes to research assistant professor Dr. David Misek, his advice and insight about the ovarian serous carcinoma samples and his underlying biology has been invaluable. Also, I would like to thank my committee members, Professor Zhan Chen, Professor Kristina Hakansson and Professor Robert Zand, for serving on my dissertation committee and for their helpful advice.

I'd also like to thank many collaborators for the research, which include Dr. Kathleen Cho and Dr. Rong Wu from department of pathololgy, the university of Michigan medical school and Dr. Kerby A. Shedden from department of statistics, The University of Michigan. I appreciate their help about the samples and data analysis for me and their patience with the ovarian cancer project over the years. I would also like to thank Dr. Timothy Barder and Eprogen, Inc.. They were gracious enough to provide the non-porous silica RP-HPLC and chromatofocusing columns upon which so much of this work depended. This work also needs to thank Dr. Steve Parus' expertise in instrumentation and software programming.

I would also like to thank members of Dr. Lubman's group, my research experience would not be enjoyable without friendship and support from them, such as Dr. Kan Zhu, Dr. Suping Zheng, Dr. Yi Zhu, Dr. Jia Zhao, Yinghua Qiu, especially Tasneem patwa, for her being such a good partner and friend in my work.

Most importantly, I would like to express my appreciation to my family, my husband Yong Chen, my daughters Erica and Michelle. Without their support and sacrifices, this work may have never been finished.

Table of Contents

List of Figures

Figure

List of Tables

Abbreviations

2-D	two-dimensional
OSE	ovarian surface epithelium RP
	reversed-phase
CF	chromatofocusing
NPS	nonporous
SB	start buffer
EB	elute buffer
HPCF	high performance chromatofocusing;
PMF	peptide mapping fingerprint.

Chapter 1

Introduction to the 2-D Liquid Phase Protein Profiling Method Combined with Mass Spectrometry and Its Application to Studies of Cancer

1.1 Introduction to Proteomics and Technologies

Cancer is a group of diseases in which cells are aggressive, invasive, and/or metastatic. Cancer remains a major public health challenge although there has been progress in detection and therapy. A substantial proportion of the US population will develop cancer during their lifetime (1) where many will die from the disease every year. In many of these cases, the cancer is not diagnosed and treated until cancer cells have already invaded surrounding tissues and metastasized throughout the body as shown in Figure 1. Many patients with breast, lung, colon and ovarian cancer have hidden or observable metastatic colonies and most conventional therapeutics are limited in their success once a tumor has spread beyond the tissue of origin. Current or future treatment strategies will have a higher probability of truly curing the disease if cancers can be detected when they are at their earliest stages, preferably in the premalignant state.

Advances in the fields of genomics and proteomics have the capability to provide insight into the molecular complexity of the disease process and thus enable the development of tools to aid in treatment and also in detection and prevention when cancer is still at their early stage. Genomics technologies have made it possible to rapidly screen for global and specific changes in gene expression that occur only in cancer cells. Genomics-based approaches include the measurements of full sets of expression of

mRNA, such as differential display (2, 3), serial analysis of gene expression (4, 5), and large-scale gene expression arrays (6, 7). Although studies of differential mRNA expression are informative, they do not always correlate with protein concentrations (8, 9), where many changes in gene expression might not be reflected at the level of protein expression or function. Proteins are often subject to proteolytic cleavage or posttranslational modifications, such as phosphorylation or glycosylation. The field of proteomics represents a new paradigm of “reverse genetics” and enables the design of more informed and interpretable studies.

Proteomics studies are the complete set of proteins found in a given cell type in any particular state. Proteomics targets the protein environment or proteome of a population of cells and can identify the changes that occur during a disease process at the protein level. The discipline of proteomics focuses on the large-scale identification and characterization of proteins through the measurement of protein expression under specific conditions from patient samples. The important tools critical to detection, diagnosis, treatment, monitoring, and prognosis in proteomics are the classification of molecules and the identification of biomarkers. Proteomic analyses of cancers by using classification and biomarker discovery techniques have provided new insights into the changes that occur in the early phases of tumorigenesis and represent a new resource for early-stage disease detection. They can provide an early indication of the disease, to monitor disease progression, to provide ease of detection, and to provide a factor measurable across populations. Different combinations of methods have been developed.

For many years, 2-dimensional polyacrylamide gel electrophoresis (2-D PAGE) followed by protein identification using MS has been the primary technique for

proteomic-based studies (10-12). Proteins are separated first according to their isoelectric point and then submitted to a second separation based upon their mass (6, 13, 14). 2-D PAGE is still the method of choice for high-resolution protein separation and provides the best method to monitor protein modifications. However, the technique suffers from a number of shortcomings, such as limited loading capacity, inability to separate proteins with extreme pI values, difficulty generating reproducible gels, or the difficulty of resolving several classes of proteins, such as very basic proteins, small proteins, and hydrophobic proteins; therefore, only part of the proteome is uncovered by the 2-D PAGE approach.

Currently, non-gel based technologies have been widely used for proteome analysis for cancer studies. Surface-enhanced laser desorption ionization (SELDI), is a technique based on the identification of proteome patterns that could be used as disease signatures. SELDI technology is based on the selective retention of proteins on modified array surfaces (15-17). When unbound proteins are discarded, proteins retained on the array are analyzed by MS, generating a specific pattern or profile of the analyzed proteome. Proteome patterns then are compared to decipher statistically significant differences in protein expression among different samples. SELDI is useful in high-throughput proteomic fingerprinting of cell lysates and body fluids.

Another approach for proteome analysis is protein arrays. Protein biochips are emerging in different formats. Several groups have been using high-throughput chip-based analysis to study protein–protein, protein–DNA, and protein–RNA interactions (18-20). Typically, a specific protein is spotted through cross-linking to a chip surface, such as a glass slide, in a grid-like fashion, and samples are passed over the slide for

detection of interacting molecules. Another emerging protein microarray format involves the surface capture of proteins through antibodies, followed by MS (21). This technique allows one to semiquantitate the expression of a protein among patients or during the course of a disease. The mass information generates a fingerprint that can be compared in a database. High-throughput can be obtained because potentially thousands of addressable locations can be designed for a chip for protein capture.

Finally, an important approach is the multidimensional protein identification technology which is based on gel-free separation (22-25). This technique combines separation of the generated proteins by liquid chromatography, enzymatic digestion of complex protein mixtures, and MS analysis. The methodology not only has greater reproducibility, but also has the ability to identify and quantify proteins, and the capability to compare results amongst different sets of experiments and samples. A major advantage of this technique is the detection of lower abundance proteins.

Our laboratory has introduced and developed a useful strategy by combining multidimensional separation of the generated proteins by liquid chromatography with mass spectrometry to study changes in cancer cells. By this strategy, we can apply the differential protein mapping method to classify cancers and find the potential biomarkers for different groups of cancers. Chromatofocusing (CF), which fractionates proteins based upon pI in the first dimension, is used with nonporous silica (NPS) reversed-phase high-performance liquid chromatography (FP-HPLC) separation in the second dimension to achieve high resolution separation. Proteins are separated and mapped to generate an image of the cellular protein content. After the two dimensional separations, there are couple of different ways to apply the differential protein method in our lab. Firstly,

classification of the cancers and differentially expressed proteins can be obtained by analyzing the data based on the UV maps. Second, after the first dimension CF separation, RP-HPLC separation can subsequently be directly interfaced to ESI-TOF-MS to generate intact molecular weight information to create a map of the accurate MW versus protein pI. This mass mapping technique can also obtain differentially expressed protein information and help to apply the classification of the cancers. Third, we can also use the two dimensional separation fractions to produce protein microarrays, where we can obtain the information about classification and biomarkers. A key point of the differential protein mapping method is the production of substantial amounts of highly purified proteins isolated in the liquid phase, which provides an automated way to collect hundreds of proteins for subsequent enzymatic digestion and identification by mass spectrometry. The automation of the method allows reproducible comparison of many samples based upon separations monitored by UV detection, allowing direct comparison of samples to detect changes in the proteome. Relevant fractions can then be subjected to MS to identify the proteins of interest. The use of differential analysis limits the number of proteins that might require further analysis by mass spectrometry techniques.

Mass spectrometry is the most important proteomic tool. From 2D-PAGE to gel-free proteomics, mass spectrometry is universally used as the end technique for the identification of proteins by peptide mass mapping and by the sequencing of peptides using tandem mass spectrometry. The highly accurate mass measurement of peptides and proteins by mass spectrometry has revolutionized the field of proteomics. In our studies here, we used several different mass spectrometers, such as ESI-TOF-MS, MALDI-TOF-

MS, QIT-TOF-MS, and LTQ. In the present study, we will describe some of the methods and experimental procedures used to accomplish this work.

1.2 Separation Methods

1.2.1 First Dimension Chromatofocusing(CF)

Chromatofocusing (CF) is a variant of ion exchange chromatography. It elutes ion exchangers by the mechanism of pH. In anion CF, proteins are bound to an anion exchanger at high pH. As the pH on the column decreases, protein positive charge becomes stronger and protein negative charge becomes weaker. pH conditions in the column eventually reach a point where a given protein's net interaction with the column becomes zero and it elutes. This anion exchange column based separation method has been selected and applied in our lab for the first separation dimension. Figure 1B shows a representation of a chromatogram. Proteins are loaded on the column at a high pH and as the titration proceeds, proteins with pI values greater than the pH will elute. Fractions are collected at a specified interval based on pH measurements. This pH-based protein separation method can rapidly fractionate large numbers of proteins and achieve separation of proteins in narrow (0.1) pH fractions. The use of online pH measurement and the column-based technique make it readily automated and easily interfaced with other liquid separations.

1.2.2 Second Dimension Non-porous silica reverse phase HPLC

After the first dimension separation, each pI fraction from a whole cell lysate still contain a large number of proteins. Reverse phase HPLC was applied to the fractions as the second dimension to separate proteins to obtain sufficient resolution. Reverse phase HPLC is one of the most widely applied analytical techniques in the world which occurs

via a partitioning of analytes between a polar, liquid mobile phase and a non-polar stationary phase. Proteins are eluted using a gradient method according to their subtle differences in hydrophobicity. Figure 1 C is a representative RP-HPLC chromatogram. This NP RP-HPLC provides rapid and highly reproducible separation of proteins. The use of C_{18}-coated, silica-based NP packing materials and small particle size (1.5µm) for the column results in improved resolution, improved protein recovery and highly efficient protein separations (26-28).

Beckman-Coulter has developed a system, the ProteomeLab PF2D which features separation by chromatofocusing in the first step, followed by RP chromatography in the second dimension. The system is shown in Figure 1A. The system is automated and allows separation of complex protein mixtures into several hundred fractions. Separation is followed by UV detection, allowing direct comparison of samples to detect changes in the proteome. In our work here, we mainly used this system. Data acquisition and fraction collection are controlled by an accessory PC. Relevant fractions can then be subjected to MS to identify the proteins of interest.

1.3 Mass Spectrometry

1.3.1 ESI-TOF-MS

ESI is a popular ionization method for peptides and proteins primarily owing to its superior ability to interface with a multitude of liquid separation techniques from LC to capillary electrophoresis. In ESI, liquid is usually sprayed through a heated, high-voltage capillary. Analyte is pumped through a capillary to form a “taylor cone”. The droplets expand through gas at atmospheric pressure, where the droplet undergoes size reduction by evaporation of solvent, so that charge density at the droplet surface increases.

When sufficient deformation and charge density electrostatic repulsion exceed the surface tension, the droplet becomes unstable and falls apart. It predominantly creates multiply charged analyte ions. The feature allows large analytes such as proteins to be measured in a mass analyzer within a limited mass range. Current ESI-TOF instruments permit the practical measurement of proteins up to 80 kDa with 100-400 ppm mass accuracy by surveying only a mass range from 0-5000 m/z.

In this work, the Micromass LCT ESI-TOF-MS instrument was used to obtain mass and quantitation data on intact proteins. Figure 3a shows the scheme of the instrument design. The eluent from HPLC is readily interfaced with ESI-TOF-MS for analysis of intact protein molecular weight. Quantitative interpretation of multiply charged ESI data is carried out using maximum entropy deconvolution (MaxEnt) software. This algorithm iteratively compares and minimizes differences between a "mock spectrum" and the original m/z data generating a parent spectrum (Figure 3b) that is most likely to result in the original m/z spectrum. Mass and intensity calculated in this manner can then be used for quantitative comparisons (Figure 3c). On-line separations with ESI-TOF-MS detection are described in greater detail in Chapters 2&3.

1.3.2 MALDI-TOF-MS

MALDI is commonly coupled with time-of-flight (TOF) mass analyzers. It is a very simple and robust method for protein identification. A protein or peptide sample is co-crystallized with an excess of a UV-absorbing matrix. Predominantly singly-charged analyte ions are created when the protein molecule mixtures are bombarded with intense, short duration bursts of laser radiation. MALDI mass spectra are easy to interpret because of their exclusively singly-charged ions. MALDI also shows a high degree of

compatibility with common biological buffers and contaminants minimizing the need for excessive sample cleanup prior to analysis. Furthermore, MALDI instruments are amenable to high-throughput approaches and spectra are generated with little sample consumption. MALDI-TOF instruments are capable of sensitivity as well as less than 10 ppm mass accuracy with internal calibration for analysis of peptides. These characteristics of MALDI-MS have been applied to its use in peptide mass fingerprinting expression for protein identification.

The Micromass TofSpec-2E is a compact high performance TOF mass spectrometer utilizing MALDI, which provides full mass spectra with high sensitivity over very wide mass ranges. Figure 4a shows the scheme of the TofSpec design. The Micromass Masslynx software package provides automatic instrument control as well as advanced data processing facilities. The mass fingerprint is searched against a protein database obtained by theoretical cleavage of protein sequences stored in databases. The protein solution eluting from the HPLC separation can be collected and each fraction is subsequently digested by trypsin. The digest from each protein peak can be further analyzed by MALDI-TOF-MS for protein identification based on peptide mass fingerprinting as shown in Figure 4b. Work details are discussed in Chapters 2&3.

1.3.3 QIT MALDI quadrupole ion trap-ToF

In our work, we used the Shimadzu AXIMA-QIT system as a method for protein identification. The AXIMA-QIT is a hybrid quadrupole ion trap - reflectron time-of-flight instrument incorporating a matrix-assisted laser-desorption ionization (MALDI) source. MS^n allows for multiple rounds of parent-ion isolation and fragmentation, making it ideal for detailed structural characterization of biomolecules, particularly oligosaccharides.

Samples mixed with a matrix are deposited onto the MALDI target and dried. A Nitrogen gas laser produces a pulse of UV light of <5ns duration at 337nm and is focussed to a spot less than 150μm diameter onto the target. The sample/matrix mixture absorbs the UV light resulting in a pulse of ions being ejected from the sample surface. These ions are accelerated and focussed by a series of electrostatic ion optics towards the ion trap. When all ions of interest are in the trap, appropriate electronics are activated and maintain the required RF voltage applied to the ring electrode. When ions are ejected from the trap, they enter a flight tube that is floated at the same potential as the extraction electrode. The energy spread is compensated for by an ion reflectron. Accuracy *is* dependent on the stability of extraction, floating and reflectron voltages and possible jitter on timing various events. Mass resolution is a function of energy spread which in turn is a function of cloud size and the same mass ions at the front and the back of the cloud will acquire different energies. The ion reflectron corrects the time-of-flight difference arising from this aberration. Figure 5a shows an overview of the system. Representative data is shown in Figure 5b. Work details are discussed in Chapter 4.

1.3.4 LC-MS/MS

We used the Finnigan™ LTQ™ system in our studies. The ThermoElectron Finnigan LTQ is a linear ion trap mass spectrometer equipped with an electrospray ionization source. A Michrom nano-HPLC system is interfaced to the mass spectrometer for automated LC-MS/MS analyses. This HPLC interface allows for separation of complex mixtures prior to on-line MS^n analysis. The Finnigan LTQ MS detector is an advanced analytical instrument that includes a syringe pump, a divert/inject valve, an atmospheric pressure ionization (API) source, an MS detector, and the Xcalibur data

system. In a typical analysis, a sample can be introduced using a valve and an LC system fitted with a column (LC/MS). In analysis by LC/MS, a sample is injected onto an LC column. The sample is then separated into its various components. The components elute from the LC column and pass into the MS detector where they are analyzed. The data from the MS detector is then stored and processed by the Xcalibur data system. With ESI, the range of molecular masses that can be analyzed by the Finnigan LTQ MS detector is greater than 100,000 u, due to multiple charging.

The LTQ has improved capacity and advantages such as improved trapping efficiency and scan speed when compared to a 3-dimensional trap. In addition, these features result in sensitivity increases that are approximately 100-fold higher than other mass spectrometers. The LTQ can be set up to collect full scan MS and MS^n (n=2-10) data. Advanced scanning modes include data-dependent scans, zoom scans, and ultra zoom scans. Ion mapping experiments can be used to generate full scan, neutral loss, and parent ion maps. Both positive and negative ions can be detected. Figure 6a shows the scheme of the system overview. Work details are discussed in Chapters 4&5&6.

1.4 Overview of the Dissertation

The work presented in this dissertation is focused on the interlysate studies of human ovarian serous carcinoma samples and breast cancer samples using differential mapping technique combined with mass spectrometry. Chapter 2 describes the initial study of 18 cell lines using the differential protein mapping method by UV map. This automated 2-D liquid fractionation system can classify the different samples according to their protein expression profiles. Potential marker bands used to classify subtypes of cancer can be identified using ESI-TOF-MS and MALDI-TOF-MS. As an outgrowth of

this initial study, the mass mapping technique was also applied to the ovarian serous carcinoma cell lines in Chapter 3. In Chapter 4, a more-in-depth profile of global protein expression patterns in 19 ovarian serous carcinoma tissue samples is presented. Molecular classification was applied and proteins that were differentially expressed in different groups were selected for identification by MALDI-TOF-MS or MALDI QIT TOF or LTQ MS. Studies were also applied to the membrane glyco-proteome using different breast cancer cell lines CA1a and AT1 in Chapter 5, which is a complimentary study for the soluble protein studies in chapter 2, 3 and 4. Multiple lectin extraction was used and differential proteins were identified using ESI-MS/MS. Chapter 6 describes the study of the classification of the humoral response using a two-dimensional liquid separation technique combined with a modified microarray method.

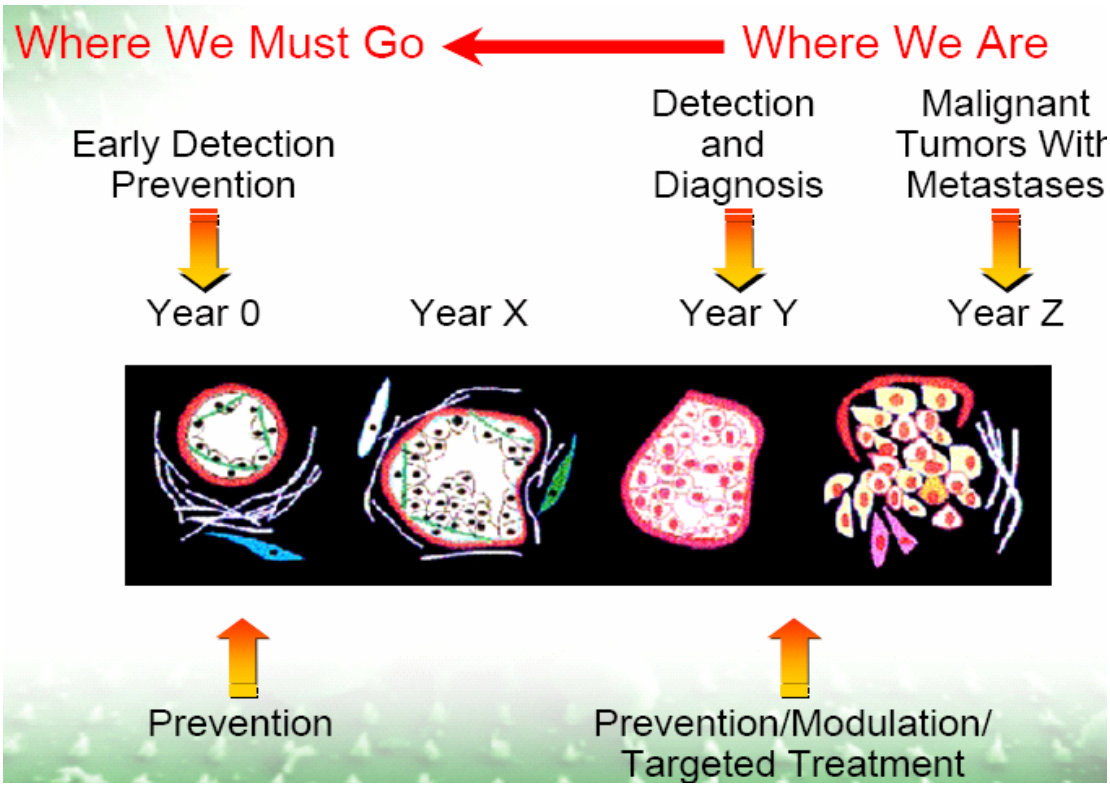
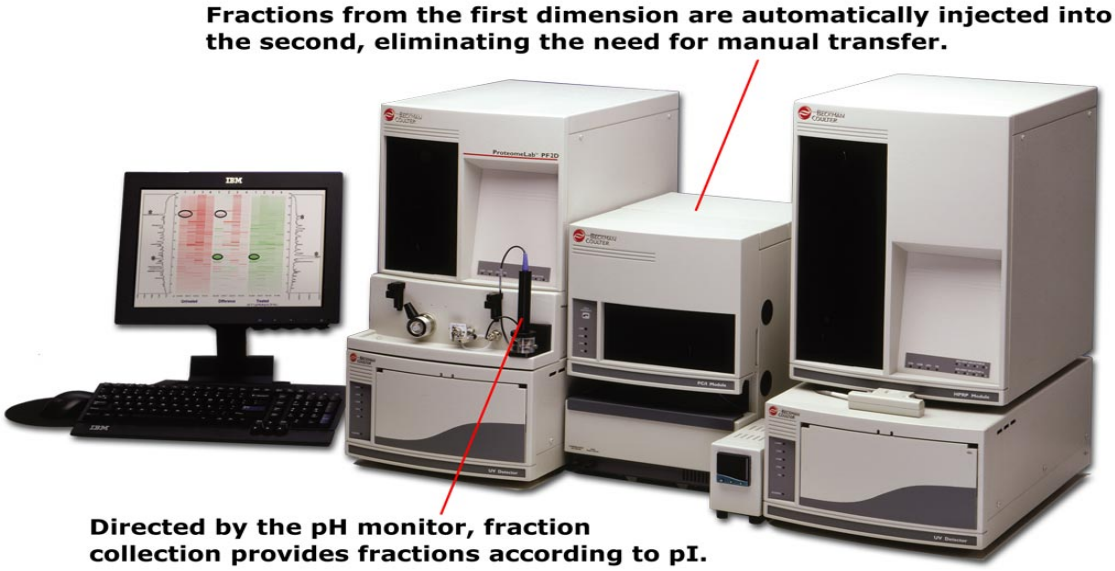
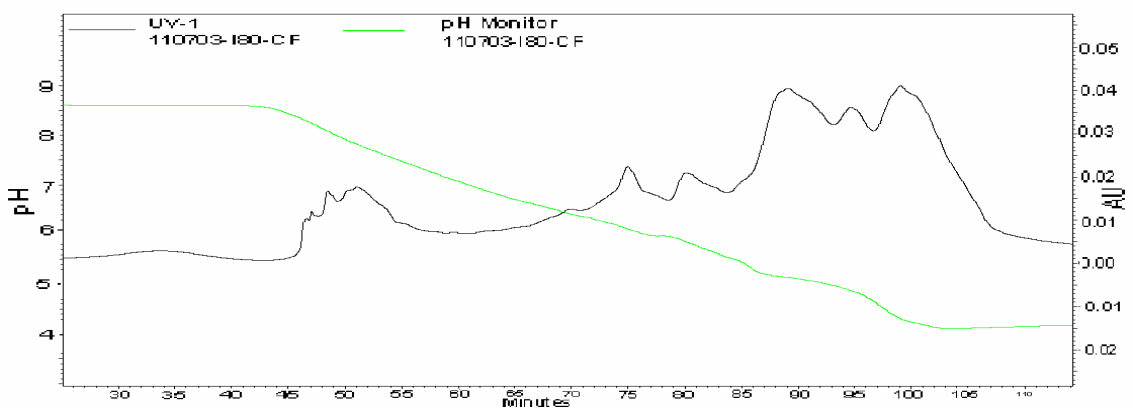


Figure 1.1 Importance of early detection of cancer

A



B



C

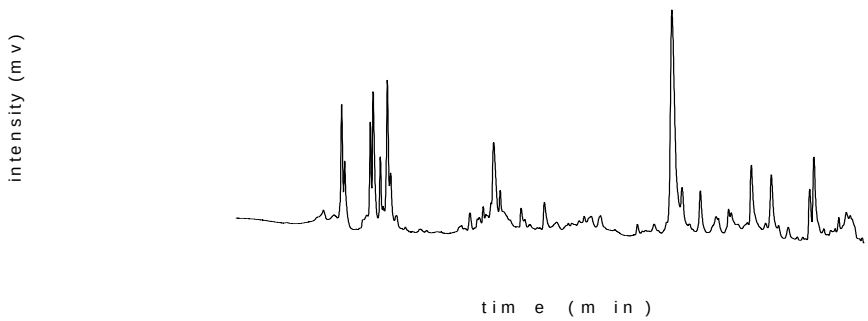
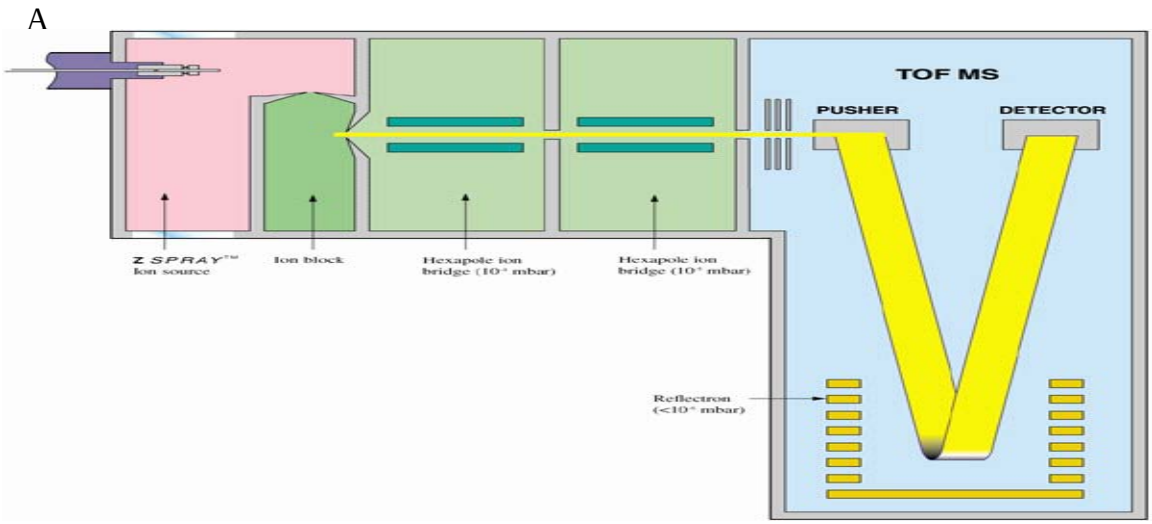
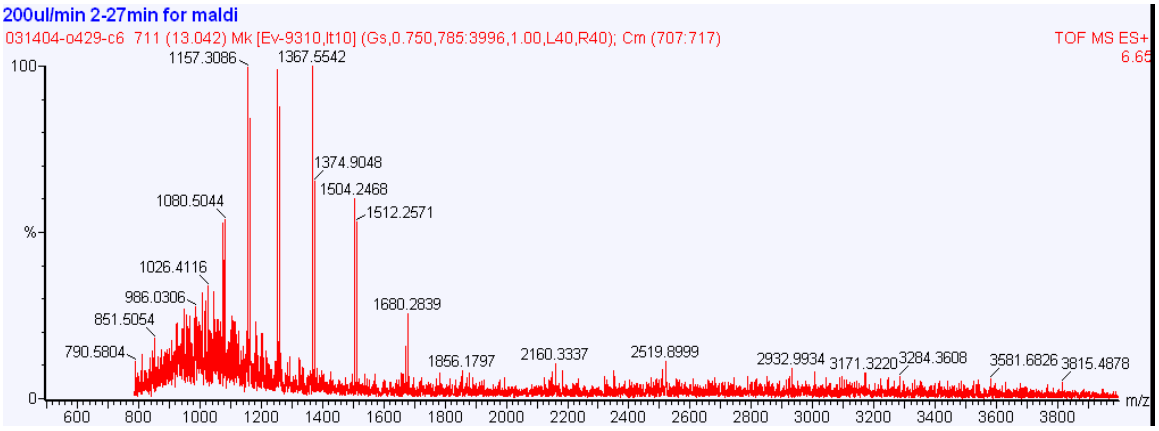


Figure 1.2 A. PF2D system
B Representative 1st dimension chromatofocusing chromatogram: UV 280nm
C Representative 2nd dimension HPLC chromatogram**:** UV 214nm

A



B



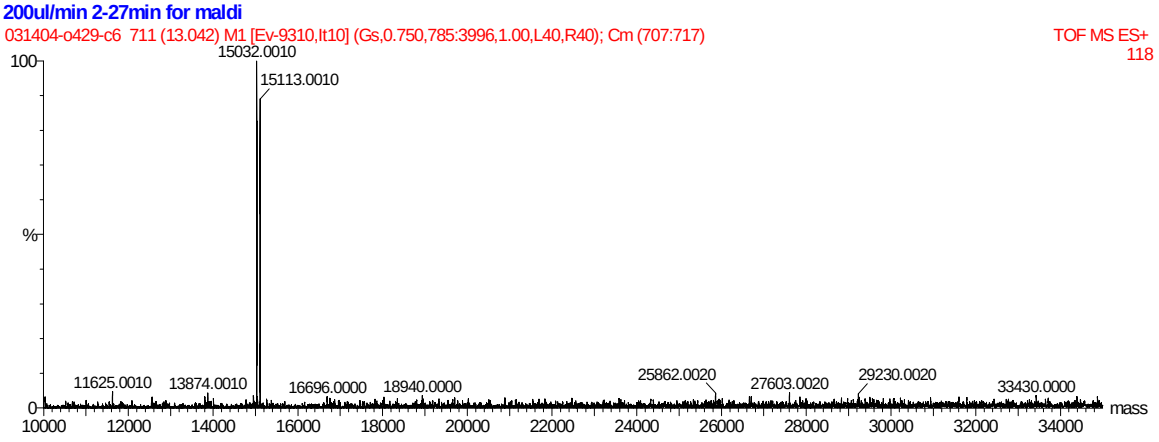
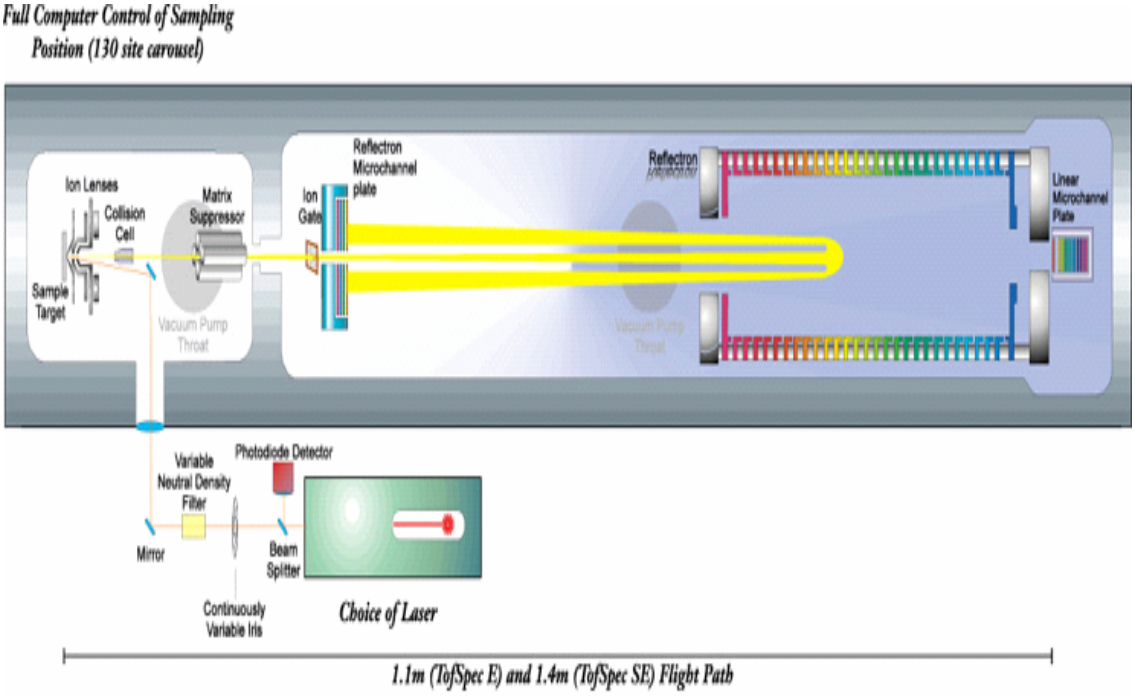


Figure 1.3 A. Schematic of Micromass LCT
B. Sample electrospray spectrum
C. MaxEnt deconvolution of mass spectrum

A



B

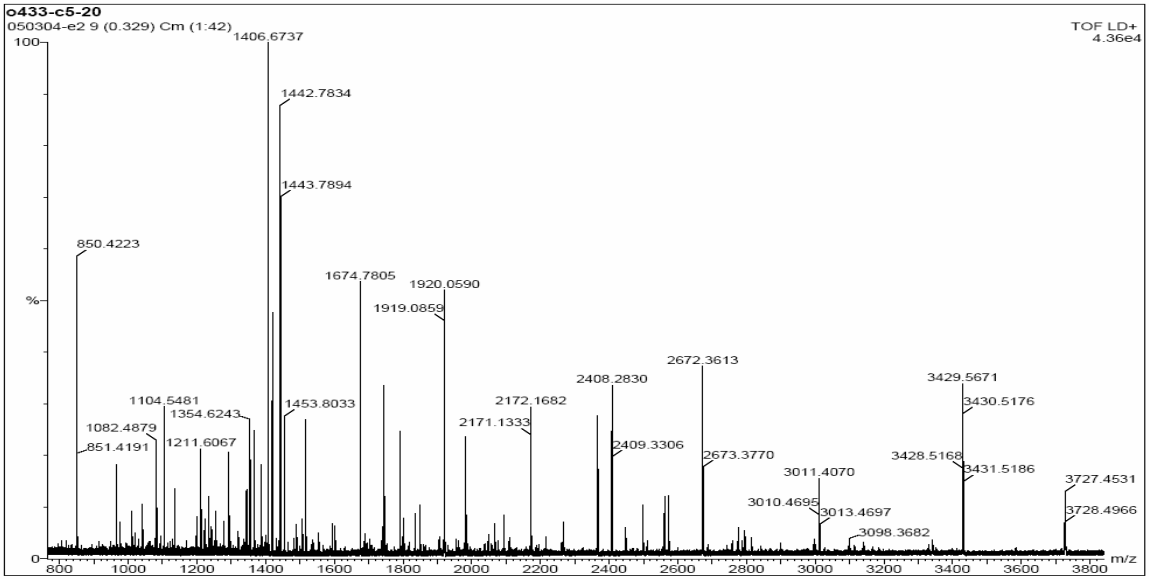
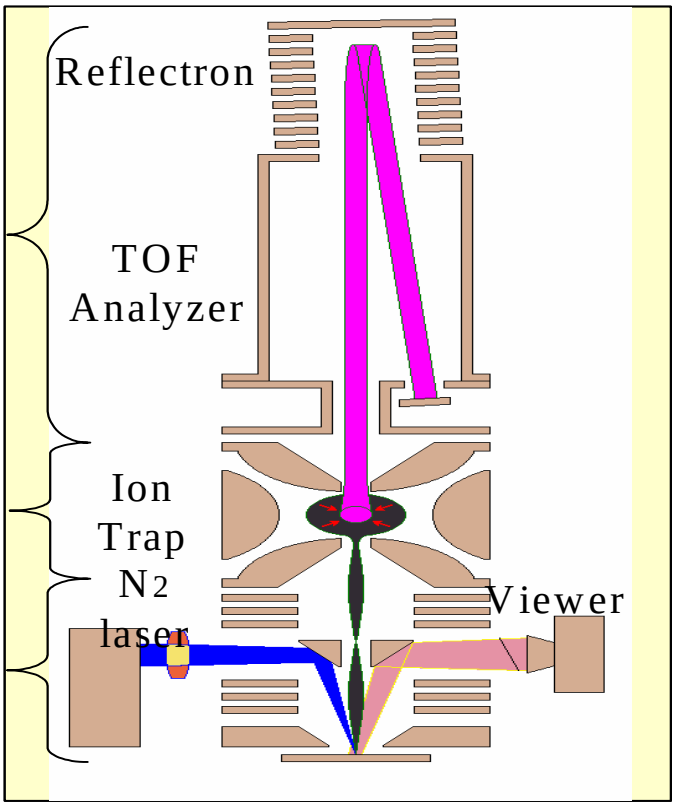


Figure 1.4 A Schematic of the Micromass TofSpec2E
B Example MALDI-TOF-MS spectrum

A



B

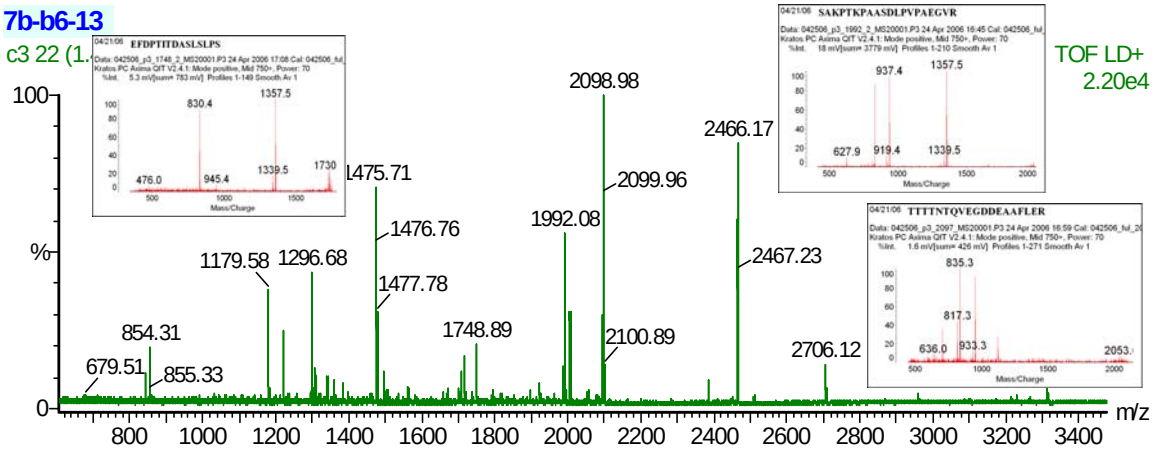
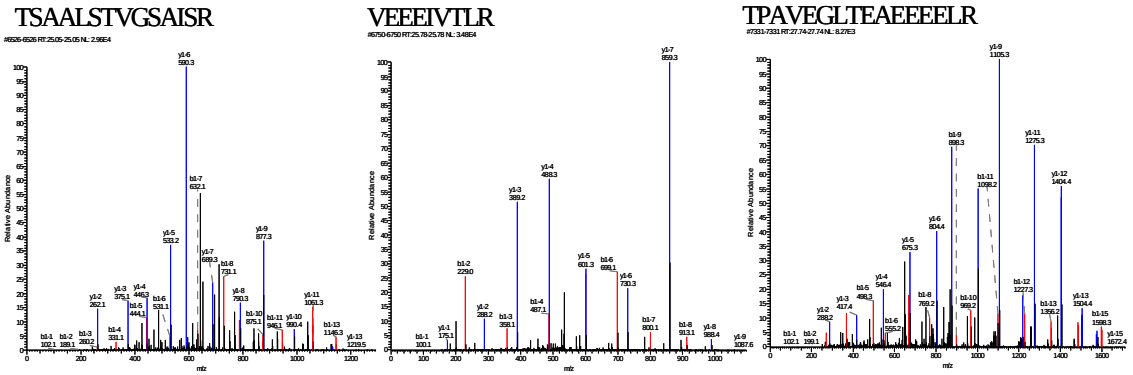


Figure 1.5 A Schematic of the Shimadzu AXIMA-QIT TOF MS
B Example MALDI QIT-TOF-MS spectrum

A

Figure 1-1. Finnigan LTQ system



MDSAGQDINLNSPNK**GLLSDSMTDVPVDTGVAARTPAVEGLTEAEEEELR**AELTK**VEEEIVTLR**QVLAAKERHCGELKRR**LGLST LGELK**QNLSRSWHDVQVSSAYVKTSEKLGEWNEKV**TQSDLYKKTQETLSQAGQKTSAALSTVGSAISR**KLGDMRNSATFKSFEDR VGTIKSKVVGDRENGSDNLPSSAGSGDKPLSDPAPF

Figure 1.6 A Finnigan™ LTQ™ MS system
B Example LTQ MS spectrum

1.5 References:

1. Chaurand, P., DaGue, B. B., Pearsall, R. S., Threadgill, D. W., and Caprioli, R. M. (2001) Profiling proteins from azoxymethane-induced colon tumors at the molecular level by matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* 1, 1320-1326.
2. Liang, P., and Pardee, A. B. (1992) Differential Display of Eukaryotic Messenger-Rna by Means of the Polymerase Chain-Reaction. *Science* 257, 967-971.
3. Douglass, J., McKinzie, A. A., and Couceyro, P. (1995) Pcr Differential Display Identifies a Rat-Brain Messenger-Rna That Is Transcriptionally Regulated by Cocaine and Amphetamine. *Journal of Neuroscience* 15, 2471-2481.
4. Ylaherttuala, S., Lipton, B. A., Rosenfeld, M. E., Sarkioja, T., Yoshimura, T., Leonard, E. J., Witztum, J. L., and Steinberg, D. (1991) Expression of Monocyte Chemoattractant Protein-1 in Macrophage-Rich Areas of Human and Rabbit Atherosclerotic Lesions. *Proceedings of the National Academy of Sciences of the United States of America* 88, 5252-5256.
5. Kal, A. J., van Zonneveld, A. J., Benes, V., van den Berg, M., Koerkamp, M. G., Albermann, K., Strack, N., Ruijter, J. M., Richter, A., Dujon, B., Ansorge, W., and Tabak, H. F. (1999) Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Molecular Biology of the Cell* 10, 1859-1872.
6. Bertucci, F., Bernard, K., Loriod, B., Chang, Y. C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K., and Jordan, B. R. (1999) Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Human Molecular Genetics* 8, 1715-1722.
7. McRedmond, J. P., Park, S. D., Reilly, D. F., Coppinger, J. A., Maguire, P. B., Shields, D. C., and Fitzgerald, D. J. (2004) Integration of proteomics and genomics in platelets - A profile of platelet proteins and platelet-specific genes. *Molecular & Cellular Proteomics* 3, 133-144.
8. Anderson, L., and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533-537.
9. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology* 19, 1720-1730.
10. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 1, 377-396.
11. Yan, J. X., Wait, R., Berkelman, T., Harry, R. A., Westbrook, J. A., Wheeler, C. H., and Dunn, M. J. (2000) A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. *Electrophoresis* 21, 3666-3672.
12. Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., and Yates, J. R. (1998) Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography microspray and nanospray mass spectrometry. *Analytical Biochemistry* 263, 93-101.
13. Cordwell, S. J., Wilkins, M. R., Cerpapoljak, A., Gooley, A. A., Duncan, M., Williams, K. L., and Humpherysmith, I. (1995) Cross-Species Identification of Proteins Separated by 2-

Dimensional Gel-Electrophoresis Using Matrix-Assisted Laser-Desorption Ionization Time-of-Flight Mass-Spectrometry and Amino-Acid-Composition. *Electrophoresis* 16, 438-443.

14. Lahm, H. W., and Langen, H. (2000) Mass spectrometry: A tool for the identification of proteins separated by gels. *Electrophoresis* 21, 2105-2114.
15. Li, J. N., Zhang, Z., Rosenzweig, J., Wang, Y. Y., and Chan, D. W. (2002) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clinical Chemistry* 48, 1296-1304.
16. Issaq, H. J., Veenstra, T. D., Conrads, T. P., and Felschow, D. (2002) The SELDI-TOF MS approach to proteomics: Protein profiling and biomarker identification. *Biochemical and Biophysical Research Communications* 292, 587-592.
17. Merchant, M., and Weinberger, S. R. (2000) Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21, 1164-1177.
18. Yan, H., Park, S. H., Finkelstein, G., Reif, J. H., and LaBean, T. H. (2003) DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science* 301, 1882-1884.
19. Lee, K. B., Park, S. J., Mirkin, C. A., Smith, J. C., and Mrksich, M. (2002) Protein nanoarrays generated by dip-pen nanolithography. *Science* 295, 1702-1705.
20. Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000) Analysis of yeast protein kinases using protein chips. *Nature Genetics* 26, 283-289.
21. de Wildt, R. M. T., Mundy, C. R., Gorick, B. D., and Tomlinson, I. M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nature Biotechnology* 18, 989-994.
22. Lubman, D. M., Kachman, M. T., Wang, H. X., Gong, S. Y., Yan, F., Hamler, R. L., O'Neil, K. A., Zhu, K., Buchanan, N. S., and Barder, T. J. (2002) Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 782, 183-196.
23. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* 19, 242-247.
24. Lustig, D. A., and Lubman, D. M. (1991) A Continuous-Flow Probe Method for Online Introduction of Liquid Samples for Detection by Laser Desorption with Resonant 2-Photon Ionization in Supersonic Beam Mass-Spectrometry. *Review of Scientific Instruments* 62, 957-962.
25. Ito, Y., Takeuchi, T., and Ishii, D. (1985) Direct Coupling of Micro High-Performance Liquid-Chromatography with Fast Atom Bombardment Mass-Spectrometry. *Journal of Chromatography* 346, 161-166.
26. Jenke, D. R. (1996) Practical examination of a nonporous silica stationary phase for reversed-phase fast LC applications. *Journal of Chromatographic Science* 34, 362-367.
27. Barder, T. J., Wohlman, P. J., Thrall, C., and DuBois, P. D. (1997) Fast chromatography and nonporous silica. *Lc Gc-Magazine of Separation Science* 15, 918-&.
28. Wirth, H. J., Unger, K. K., and Hearn, M. T. W. (1991) High-Performance Liquid-Chromatography of Amino-Acids, Peptides and Proteins .109. Investigations on the Relation between the Ligand Density of Cibacron Blue Immobilized Porous and Nonporous Sorbents and Protein-Binding Capacities and Association Constants. *Journal of Chromatography* 550, 383-395.

Chapter 2

Classification of Cancer Cell Lines Using an Automated 2-D Liquid Mapping Method with Hierarchical Clustering Techniques

2.1 Introduction

An important area in current cancer research is directed towards improving the molecular classification of cancer (1). This involves the ability to find markers, which can be used to differentiate cancers arising in the breast, lung, colon, prostate, and other organs from one another (2, 3). Moreover, various subtypes exist for cancers arising in many organs, each with different histopathology and ultimately very different clinical outcomes (4-6). It is essential to find markers that can be used to classify tumors with respect to organ of origin, benign versus aggressive behavior, and likely response to specific treatments (7). This ability to use molecular information to improve upon existing morphology-based tumor classification schemes would further the concept of personalized medicine.

Ovarian cancer is particularly problematic in that it comprises a heterogeneous group of tumors, many with poorly characterized precursor lesions. Worldwide, ovarian cancer is the sixth most common cancer in women with the highest incidence rates appearing in developed countries (1). Epithelial ovarian cancer, or ovarian carcinoma (OvCa), constitutes about 90% of all ovarian cancers and is divided into four major distinct subtypes (serous, mucinous, endometrioid, and clear cell) based on their morphological features. Interestingly, each ovarian carcinoma subtype resembles normal

epithelial cells found elsewhere in the female reproductive tract derived from a common embryological precursor known as the coelomic mesothelium (5). For example, serous, mucinous, and endometrioid ovarian carcinomas display morphological features similar to normal epithelial cells lining the fallopian tube, endocervix, and endometrium, respectively. Presently, ovarian carcinoma is managed clinically without consideration of morphology, yet there is growing clinical-pathologic and molecular evidence that the different subtypes may represent clinically, biologically and genetically distinct disease entities.

There are several strategies that can be used to classify cancers based upon either gene or protein expression. In particular, DNA microarrays have been used to characterize global gene expression patterns of cancer samples (1, 4). This technology has enabled the study of comprehensive gene expression profiles of large numbers of tumor samples, which can be used to classify cancers based upon characteristic gene expression patterns. In recent work, for example, Schaner et al used DNA microarrays to identify groups of genes that could distinguish ovarian from breast carcinomas, clear cell subtype from other ovarian carcinomas, and grade I and II from grade III serous papillary carcinomas (4). In related work, Perou et al were able to use gene expression signatures to define subclasses of breast cancer (8) and Sorlie et al were able to correlate differences in expression patterns of breast cancers with clinical outcome and identify subclasses having poor prognosis (9). In other work, Giordano et al were able to use gene expression profiles of adenocarcinomas of the lung, colon and ovary to demonstrate the ability to classify tumors in an organ-specific manner (3). Schwartz et al used gene expression patterns to classify the different subtypes of ovarian cancer and showed that these

patterns in ovarian adenocarcinomas reflect both the morphological features as well as the biological behavior (5). Numerous other studies have also used gene expression to classify various cancers and their subtypes and their relationship to one another (10-12).

An alternative means of classifying different types of cancers involves profiling the protein expression of cells or serum (7, 13). The use of protein expression may be most informative for classification of cancers since mRNA and protein expression from a given gene may be discordant and it is ultimately the protein expression that determines the function and structure of the cells. In addition, protein expression can be profiled from either tissues or serum. The traditional method for profiling large numbers of proteins from cells is 2-D gel electrophoresis (14). A number of studies using large numbers of quantitative 2-D gels for tumor classification have been performed for bladder, breast, lung, prostate and ovarian cancers (6, 7), where, in general, benign and malignant tumors were identified by proteins that were differentially expressed and tumor stage classified by marker proteins that were up or down-regulated. Alaiya et al did extensive work on the classification of ovarian tumors where protein expression of 40 tumor samples was evaluated using 2-D gels with hierarchical cluster analysis to distinguish borderline ovarian tumors from malignant and benign tumors (6). Other work by this group using quantitative 2-D gel electrophoresis identified protein markers that could classify benign, borderline and malignant tumors (15). In recent work Jones et al examined the use of laser capture microdissection (LCM) of human ovarian epithelial cells in tissue specimens followed by 2-D gel electrophoresis to identify proteins that change between invasive and noninvasive ovarian cancers (7). These differentially

expressed proteins could potentially be used to generate markers of early detection or therapeutic targets unique to the invasive cancer.

Although 2-D gel electrophoresis has been the most widely used technique for separating large numbers of proteins, there are still drawbacks that limit its utility as a general tool for profiling of large numbers of samples. The 2-D gel method is generally a slow, manually intensive technique that can require several days to run and stain. Moreover, the reproducibility of interlysate comparisons may be limited due to varying run conditions between gels where spots may become difficult to compare and quantitation may be limited. In addition, proteins are embedded in the gel requiring manually intensive procedures to excise the spots for further analysis by mass spectrometry.

An alternative strategy to rapidly classify large numbers of different types of cancers using protein expression profiling is 2-D liquid mapping of proteins (16, 17). This technique uses chromatofocusing as the first separation parameter followed by nonporous reversed phase HPLC as the second dimension to orthogonally map large numbers of proteins based on their pI and hydrophobicity respectively. Both dimensions of analysis use standard chromatography (HPLC) equipment designed to reproducibly handle large numbers of samples in the liquid phase. It also uses UV absorption for detection so that quantitative comparisons of protein expression can be performed between samples. The method has distinct advantages in automation in that all fractions are in the liquid phase and large numbers of samples can be run and protein maps obtained for easy differential comparisons. Moreover, because proteins elute in the liquid phase, direct interface with other methods such as mass spectrometry is readily achieved.

In this work, we demonstrated the use of an automated 2-D liquid fractionation system for the liquid phase separation and mapping of the protein expression for eight serous ovarian cancer and three ovarian surface epithelium (OSE) cell lines. Hierarchical clustering analysis was used to classify the different samples according to their protein expression profiles showing that specific types of serous carcinoma cell lines tend to cluster together. Several other cell lines e.g., ovarian clear cell and endometrioid carcinoma and breast epithelial cell lines, were also fractionated and mapped by the 2-D liquid method and cluster analysis performed on a total of 18 samples. We compared our cluster analysis to results using oligonucleotide microarrays identifying some similarities and some differences in the observed clustering patterns. We could classify different types of cancers in addition to identifying potential marker bands to classify different subtypes of individual cancer using this methodology. The use of this method also limits the number of potential marker bands that may need to be identified by mass spectrometry. We also demonstrated that these samples could be run reproducibly and in an automated fashion using this method.

2.2 Experimental Section

2.2.1 Cell Culture and Sample Preparation

A total of fourteen ovarian cell lines were applied in this study, which included eight serous, one clear cell, two endometrioid and three immortalized ovarian surface epithelial cell lines. Ovarian serous carcinoma-derived cell lines HOC-1 and HEY were a gift from L. Dubeau (USC School of Medicine, Los Angeles, CA); DOV13, OVCA420, OVCA429, OVCA432, and OVCA433 were a gift of D. Fishman (North-western University, Chicago, IL); PEO1 was a gift from T. Hamilton (Fox Chase Cancer center,

Philadelphia, PA). One ovarian clear cell adenocarcinoma derived cell line (ES-2) and two ovarian endometrioid adenocarcinoma cell lines (MDAH-2774 and TOV-112D) were obtained from the American Type Culture Collection. Ovarian surface epithelial cells expressing SV40 large T Antigen (IOSE-80 and IOSE-144) were a gift of N.Auersperg (University of British Columbia, Vancouver, British Columbia, Canada). HOSE-A (96.9.18), an OSE cell line immortalized with HPV16 E6/E7 was a gift from W. Lancaster (Wayne State University School of Medicine, Detroit, Michigan). The MCF10 human breast epithelial cell lines were obtained from the Barbara Ann Karmanos Cancer Institute (Wayne State University, Detroit, MI, USA). Cells were cultured as previously described (18). All the cell lines were cultured under standard conditions in the Department of Pathology at the University of Michigan Medical School. The cells were maintained in DMEM/10% fetal bovine serum and 1% Penicillin/Streptomycin (Invitrogen, Gaithersburg, MD). When monolayer cultures were 80-90% confluent, the cells were washed three times with phosphate buffered saline (PBS) and lysed in lysis buffer. Lysis buffer (90 μL/cm^2) consisted of 6 M urea (INC Biochemicals, Cleveland, OH, USA), 2M thiourea (INC), 1% n-octyl-β-D-glucopyranoside (OG1) (Sigma, St. Louis, MO, USA), 2mM DTT (Sigma), 2% Biolyte ampholytes pH 3-10 (Bio-Rad, Richmond, CA, USA), and 2.5 mM PMSF (Bio-Rad). The cells were removed by scraping with a cell scraper (Costar, Cambridge, MA, USA). The insoluble material was removed by ultracentrifugation at 35000 rpm at 4 ºC for 1h, and the supernatant was stored at -80ºC for further use.

2.2.2 Chromatofocusing(CF) and Nonporous (NPS) RP-HPLC Separation of Ovarian Serous Carcinoma Cell Line Lysates

CF and NPS RP–HPLC were performed continuously using an integrated protein fractionation system ProteomeLab™ PF 2D (Beckman Coulter, Inc. Fullerton, CA, USA). An HPCF-1D column (250 × 2.1 mm) was used to perform chromatofocusing. Two buffers, a start buffer (SB) (Beckman Coulter, Inc. Fullerton, CA, USA) and an elution buffer (EB) (Beckman Coulter, Inc. Fullerton, CA, USA), were used to generate the pH gradient on the column. Both buffers were prepared in 6 M urea and 0.2% octyl glucoside. Before running the CF, the pH of SB was adjusted to 8.5 +/- 0.1 and EB was adjusted to 4.0+/- 0.1 using either a saturated solution (50 mg/mL) of iminodiacetic acid (Sigma, part # I5629) if the buffer was too basic or 1 M NH_4OH if the buffer was too acidic. A PD-10 G-25 column (Amersham Pharmcia Biotech) was used to exchange the protein sample from the lysis buffer to the equilibration buffer used in the CF experiment.

The HPCF 1D column was first flushed with 100% distilled water (filter through a 0.45 um filter) for 10 column volumes at 0.2 mL/min, then equilibrated with 100% SB for 30 column volumes. After equilibration with SB, the HPCF column was ready to start the ProteomeLab PF 2D default method where injection of the sample began the method. After the method had been started, the column was washed with 100% SB to remove material that did not bind to the column at pH 8.5. When the wash was complete, the UV absorbance returned to baseline. Once a stable baseline was achieved, the method was initiated at 100% EB. UV detection was performed at 280 nm and the pH was monitored on-line by a flow-through pH probe (Beckman Coulter, Inc. Fullerton, CA, USA). As the pH decreased, pH fractions were then collected in 0.15 pH intervals where 30 fractions in total were collected in the range of pH 8.5 - 4.0. After the pH of the eluent reached 4.0, the HPCF column was washed with 10 column volumes of 1M NaCl and the fractions

collected by time. After the salt wash, the HPCF column is washed with 10 column volumes of distilled or deionized water. The CF portion of the method for the ProteomeLab PF 2D required around 185 minutes.

When the first-dimension separation was completed, the pI fractions collected from the first dimension were then automatically run on the second dimension based on the specified ProteomeLab PF 2D sequence. Proteins were resolved by reversed-phase chromatography using a HPCF-2D (4.6 × 33 mm) NPS column (Beckman Coulter, Inc. Fullerton, CA, USA) and detected by absorbance at 214 nm using a Beckman model 166 UV absorption detector. Solvent A was 0.1% trifluoroacetic acid (TFA) in water with 0.05% n-octyl β–D-galactopyranoside and solvent B was 0.08% TFA in acetonitrile with 0.05% n-octyl β–D-galactopyranoside. The Gradient was run from 15 to 25% B in 1 min, 25 to 35% in 6 min, 35 to 38% in 4 min, 38 to 45% in 6 min, 45 to 65% in 2 min, 65 to 67% in 6 min, and finally up to 100% in 1 min, then back to 5% in 1 min. After the gradient, the column was washed by two fast gradients from 5% B to 100% B in 5 min, 100% B back to 5% B in 1 min. At the end of each 2^{nd}-dimension run, the method equilibrates the column with an initial mobile phase (A) for 10 column volumes. The flow rate used was 0.75 mL/min and the column temperature was 65 ºC. Proteins were collected for further analysis using an automated fraction collector. The method automatically saves the raw UV absorbance data for each 2^{nd}-dimension analysis of the chromatofocusing fractions for protein mapping and data analysis using ProteoVue^{TM} in the PF 2D Software Suite.

2.2.3 MALDI-TOF-MS Sample Preparation and Data Acquisition

Before MALDI analysis, each tryptic digested sample was desalted and concentrated using C18 ZipTip (Millipore) into 5 μL 60% (v/v) ACN/0.1% (v/v) TFA. One microliter of the concentrated peptide mixture was spotted onto a Micromass 96-spot plate followed by 1 μL of matrix-standard layered on top of it. The MALDI matrix solution was prepared by diluting saturated α-CHCA (Sigma) solution with 60% v/v ACN and 0.1% v/v TFA at 1:4 ratio v/v with three internal standard proteins: angiotensin I (1296 Da), adrenocorticotropic hormone (ACTH clip 1–17, 2093 Da), and ACTH clip 18–39 (2465 Da) (Sigma). 1 mg/mL standard proteins were diluted 100-fold with deionized water. These standards were further diluted 50, 40, and 30 times, respectively, with the diluted matrix.

The Micromass TofSpec-2E™ (Waters) was used to generate peptide mass fingerprints (PMF), followed by database searching in the SwissProt database. Spectra were calibrated using the internal calibrants to achieve a mass accuracy of less than 50 ppm. Monoisotopic peak lists were generated and searched against the SwissProt protein database, using the MS-Fit search engine (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm). The search was performed using the following parameters: (1) species: human; (2) allowing one missed cleavage; (3) possible modifications: peptide *N*-terminal glutamine to pyroglutamic acid, oxidation of methionine, and protein *N*-terminus acetylated and phosphorylation of S, T and Y; (4) peptide mass tolerance 50 ppm; (5) MW ranged from 1000 to 100,000 Da; (6) p*I* range of protein 3–10. Protein identifications were considered confident matches when meeting the following criteria: (i) ranked in the top three database hits; (ii) sequence coverage is >20%; (iii) average mass error is <50 ppm (iv) MOWSE scores over 10^4.

2.2.4 Electrospray/Ionization TOF-MS

ESI-TOF-MS analyses of collected protein peak fractions are performed on a Micromass LCT™ workstation (Waters, Milford, MA, USA). Ions are generated from a z-spray source with a desolvation gas temperature of 150°C at a flow rate of approximately 450 L/h. The source is held at 100°C. The capillary voltage is +2800 V, the sample cone is +45 V, the extraction cone is +2 V, and the hexapole RF is +750 V with a DC offset of +3 V. The second hexapole is biased to -3 V, and the detector is held at -2750 V. External calibration is performed by direct infusion of a standard NaI/CsI (Sigma) solution. Fractions are infused into the source using a syringe pump (Harvard Apparatus, Holliston, MA, USA) at a flow rate of 30–50 μL/min. Data is processed using Micromass MassLynx™ v3.4 software (Waters), and protein multiply charged umbrellas are deconvoluted using Micromass MaxEnt® 1 software (Waters). Deconvolution is performed using a mass range of 5–85 kDa, 1 Da resolution, 0.75 Da peak width, and a 65% peak height value.

2.2.5 Software

The data from the 2-D liquid separations are displayed using ProteoVue and DeltaVue software available in the PF 2D Software Suite (Beckman Coulter, Inc. Fullerton, CA, USA). The chromatographic UV intensities result from the NPS HPLC second-dimension separation of each pI fraction which were converted and displayed in a 2-D “lane and band” format by the ProteoVue software resulting in a highly detailed pI versus hydrophobicity protein expression map. ProteoVue allows comparison of multiple or all second-dimension runs for one sample in a 2-D map using either gray scale or a color-coded format where color hue or its intensity is proportional to the relative

quantitative UV intensity of each peak. Relationships or patterns within a complex chromatographic data set can be easily viewed in this format. The DeltaVue software allows side-by-side viewing of the second-dimension runs for two samples or two groups of samples so that differences in protein expression between them can be compared. This software quantitatively displays one protein map in shades of red and the other map in shades of green. The difference between the two maps is obtained by point-by-point subtraction or by area difference and displayed as a third map in the middle. The color (red or green) at a particular location in the difference map indicates which protein is more abundant, and the color brightness indicates the quantitative difference. The program also provides a means to obtain a quantitative number between the expression levels of protein in the two samples.

2.2.6 Data analysis and clustering

2.2.6.1 Data standardization:

The raw UV data for each sample were standardized to remove differences in the level and slope of the baseline. To do this, for each point, the 10^{th} percentile within a window of +/- 50 measurements was calculated and subtracted from the point. Then negative values were replaced with zero, 0.0001 was added to all values, and the data were log transformed.

2.2.6.2 Alignment:

Standardized UV data for each pair of samples were aligned in order to maximize the local correlation coefficients between the aligned samples. Specifically, suppose samples A and B are to be aligned. An alignment is defined by a sequence of index pairs $(t_a(1),t_b(1))$, $(t_a(2),t_b(2))$, ... such that $(t_a(k+1)-t_a(k), t_b(k+1)-t_b(k))$ is equal to (1,0), (1,1),

or (0,1). That is, at each step either the A sequence advances by one index, the B sequence advances by one index, or both sequences advance by one index. At the initial point either ta(1) or tb(1) is equal to 1, and at the final point either ta or tb is equal to the length of the data sequence (corresponding to the greatest measured hydrophobicity value).

To evaluate alignment quality, for each pair of indices t_a in sample A and t_b in sample B such that $|t_a - t_b| < 150$, the correlation coefficient between the data values $A(t_a-75)...A(t_a+75)$ and $B(t_b-75)..B(t_b+75)$ was calculated. The goal is to maximize the sum of local correlation coefficients over all possible t_a, t_b sequences that always remain within 150 units of each other. This problem can be efficiently solved using dynamic programming techniques.

2.2.6.3 Comparisons:

To compare the overall pattern of protein expression in the samples, each pair was aligned separately, as described above, then a correlation matrix was formed by calculating the Pearson correlation coefficient between each aligned pair of samples. These correlation matrices were then visualized using a hierarchical clustering technique. The hierarchical clustering technique produces a dendrogram in which pairs of points are joined sooner (i.e. closer to the ends of the dendrogram) if they have greater correlation. Complete linkage clustering was used to define the dendrograms.

2.2.7 Biomarker Identification:

To identify potential biomarkers, all samples were first aligned to a single sample selected as the standard. Then comparisons were made separately at each hydrophobicity level between two groups of samples. Values were selected if the ratio between mean

levels within the two groups exceeded 4, and if the t-test p-value between the two groups was less than 0.05. In addition, at least 25 consecutive hydrophobicity levels were required to meet these conditions in order for the band to be considered as a biomarker.

2.3 Results and Discussion

2.3.1 Cell lines

Eight ovarian serous carcinoma-derived cell lines and three OSE cell lines (IOSE-144, IOSE-80 and HOSE-A) were used to study ovarian cancer proteomes (Table 1). HOC-1 was derived from ascites tumors of a patient with well-differentiated serous adenocarcinoma of the ovary (19). HEY was established from a moderately differentiated papillary (serous) cystadenocarcinoma of the ovary, which had been passaged previously as a xenograft in immunocompromised mice (19). PEO1 was derived from the ascites of a patient with poorly differentiated serous adenocarcinoma after treatment by chemotherapy (20). Both HEY and PEO1 produce tumors in immunocompromised mice (19, 20), whereas HOC-1 does not. The ovarian carcinoma cell lines DOV13, OVCA420, OVCA429, OVCA432, and OVCA433 were originated from serous cystadenocarcinoma (21).

The clear cell carcinoma line ES-2 and two endometrioid carcinoma cell lines MDAH-2774 and TOV-112D were also mapped by the 2-D liquid method for comparison to the serous carcinoma cell lines. ES-2 was established from a surgical tumor specimen described as a poorly differentiated ovarian clear cell carcinoma. Tumors develop when nude mice are inoculated subcutaneously (22). The MDAH-2774 cell line was developed from cells in the ascites fluid from a patient with endometrioid ovarian cancer and forms tumors in nude mice (23). TOV-112D was derived from an

endometrioid carcinoma from a patient who were never exposed to chemotherapy or radiation therapy (24).

Three MCF10 breast epithelial cell lines [MCF10AT1, MCF10CA1a.cl1 (CA1a) and MCF10CA1d.cl1 (CA1d)] were also mapped and compared to the ovarian carcinoma cell lines. The MCF10A cell line originated from a patient with fibrocystic disease (25). MCF10AT1 cells are MCF10A cells transformed with a mutant c-Ha-ras protein (T24). This line forms preneoplastic lesions in nude mice that represent a premalignant stage with potential for neoplastic progression. CA1a and CA1d were derived from xenografts of MCF10AT1. When subcultured and re-xenografted, these lines rapidly form invasive carcinomas with metastatic potential and display histologic variations ranging from undifferentiated carcinomas to well-differentiated adenocarcinomas (26).

2.3.2 Analysis of ovarian cancer proteomes

The primary objective of this study was to classify large numbers of samples by mapping and comparing 2D liquid maps generated with UV detection so that we can obtain basic information on the similarities and differences among ovarian cancer cell lines. OSE and serous carcinoma-derived cell lines were fractionated using chromatofocusing at 0.15 pI intervals. Each of these fractions was automatically collected and continuously injected into the second dimension column using nonporous silica (NPS) RP-HPLC. The result is a virtual 2-D UV map profile, which displays the pI versus hydrophobicity of the protein expression for the whole cell lysate. Protein detection in the first dimension step is performed using UV absorption at 280 nm and 214 nm in the second dimension. As an example, the profile of serous carcinoma HOC-1 is shown in Figure 1. In Figure 1, ~4.0 mg or 2.0 $\times 10^7$ cells were loaded onto the first dimension

chromatofocusing column. In this figure, each lane corresponds to a different pI value and the bands correspond to the hydrophobicity as generated by the % acetonitrile on the HPLC gradient at that pI. For each map, a total of 28 pI fractions have been mapped which correspond to a pH range of 4.11 – 8.32. The more acidic and basic fractions are not shown here because few proteins were detected. Many of the RP-HPLC fractions obtained after chromatofocusing contained as many as 60-100 proteins. As a representative example the different proteins separated by NPS RP-HPLC from the pI fraction 4.85 -5.00 of sample DOV13 are shown in Fig. 2. It is estimated that each sample was fractionated into over 1500 protein bands using the 2-D liquid mapping method.

In each of the samples used to generate the 2-D UV map, the protein content was determined by the Bio-Rad protein assay. Equal amounts of protein were loaded onto the first dimension chromatofocusing column so that a quantitative comparison could be obtained. The UV patterns of ovarian cancer proteins were highly reproducible within each cell line. The CF reproducibility is shown in Figure 3a, where sample IOSE-144 was run by CF three separate times. In each case different amounts of sample were loaded onto the CF column, 3.5mg, 4.0mg and 4.5mg respectively. The second dimension hydrophobicity profiles reproducibility is highlighted for one of the pI lanes from the total expression profile and is shown in Figure 3b. These results show that the chromatogram of the band patterns and retention times in both the first dimension chromatofocusing separation and the second dimension reverse phase liquid separation are very similar. This was found to be the case in all the pI lanes separated for all samples (data not shown). The 2-D liquid mapping protocol allows the production of highly

reproducible differential maps where hundreds of proteins can be compared by computer analysis for a large number of samples.

The digitized protein profile that results from our liquid separation is relatively reproducible from run to run for the same sample. Reproducible 2-D liquid separations with well resolved protein peaks are a prerequisite for the establishment of a reference map of the proteome of the ovarian serous carcinoma cell lines for interlysate comparisons. The overall resolution and reproducibility pattern of the 2D separation was greatly improved by the automated system. This reproducibility makes it possible to compare different cell lines for large numbers of samples, which is essential for searching for biomarkers.

The peak patterns for each pI fraction of each cell line were aligned using a dynamic programming technique. The samples were transformed for reasonable statistical analysis. The transform first subtracts the baseline of the curve so that a flat baseline at zero was obtained; next, each point was taken as the natural log. A difference of 1 unit on the y-axis corresponds to a roughly 2.7 fold difference in the UV readout. Two original data sets of sample HEY and PEO1, fraction pH 4.55-4.70, gradient range from 41.5% to 49.5% are shown in Figure 4a. The index of data pairs aligned together is used as x-axis units and the converted intensity of each peak as we described above is used as y axis units. The data sets that were aligned using this method are shown in Figure 4b. It is seen that the two traces are not well aligned originally, but after stretching of the profile, the alignment is much improved. The chromatographic profiles are properly aligned to compensate for minor drifts in retention times. These small retention time shifts may be due to changes in the columns during use, minor changes in mobile

phase composition, drift in the instrument, interaction between analytes, etc. This alignment technique compensates for these drifts and allows comparisons of protein bands in different cell lines for large numbers of samples.

In order to compare the overall pattern of protein expression in the samples, after each pair was aligned separately, a correlation matrix was formed by calculating the Pearson correlation coefficient between each aligned pair of samples. These correlation matrices were then visualized using hierarchical clustering techniques. Here we used three different cluster analyses, "single linkage", "complete linkage", and "average linkage" hierarchical clustering (27). Differences between these three methods arise because of the different ways of defining distance (or similarity) between clusters. Single linkage is also known as the nearest neighbor technique where the distance between groups is defined as the distance between the closest pair of objects and only pairs consisting of one object from each group are considered. The complete linkage involves the farthest neighbor, where the clustering method is the opposite of single linkage--distance between groups is defined as the distance between the most distant pair of objects, one from each group. For average linkage, the distance between two clusters is defined as the average of distances between all pairs of objects, where each pair is made up of one object from each group. All the fractions, from pH 8.32 to pH 4.11, were analyzed by these three different methods. Total results are obtained by the average of all the fractions and are shown in Figure 5a, 5b and 5c. In the dendrograms, the length and the subdivision of the branches display the relatedness of the cell lines and the expression of the proteins.

The dendrograms in figure 5 shows the relationship of the different serous carcinoma cell lines based upon their protein expression using the 2-D liquid mapping technique. In order to evaluate the method, we analyzed two sets of samples as IOSE-144-1 and IOSE-144-2. It was found that using these three different clustering methods, OSE cell lines IOSE-144 and HOSE-A clustered together (Figure 5), while the two IOSE-144 samples clustered together most closely as expected. In addition, several serous carcinoma cell lines clustered with each other, i.e., DOV13, OVCA429, and OVCA433 clustered together. It is interesting that IOSE-80, which was derived from OSE, clusters with serous carcinoma lines HEY and PEO1. Indeed, IOSE-80 appears to be strongly linked to PEO1 throughout the pI fractions of the 2-D maps. This is not surprising since the IOSE-80 cell line has been cultured for many passages and may have obtained some of the characteristics of the carcinoma-derived lines. Although limited information is available on the cell lines, it is shown that the protein expression of certain cell lines is closely related to others and that these cluster together on the dendrogram. In order to show the relationship between some of these cell lines that cluster together, the corresponding protein expression maps for a given pH fraction are shown in Figure 6. In the fractions that cluster together a number of protein bands are common to the different cell lines that define this relationship.

In addition to the serous carcinoma cell lines, several other ovarian and breast tumor cell lines were mapped by the 2-D liquid method. It is interesting to note that ES-2, MDAH-2774 and TOV-112D cluster closely together. This is not unexpected given the detailed 2-D mapping of proteins in previous work (17, 28) where the ES-2 clear cell line and MDAH-2774 were found to have many bands in common between them and in more

recent work (29) where TOV-112D was found to be somewhat similar also. Other interesting clusters include those from the MCF10 breast epithelial cell lines. Ca1d and Ca1a are related highly malignant breast cancer lines derived from xenografts of AT1 and cluster together as expected. In addition, AT1, which is a premalignant breast epithelial cell line also derived from MCF10, clusters closely with the other breast cancer cell lines as expected.

In order to quantify differences between the immortalized OSE lines (IOSE-144, HOSE-A), ovarian carcinoma-derived lines (ES-2, MDAH-2774, TOV-112D), and breast epithelial (AT1, Ca1a, Ca1d) cell line groups, the pair-wise correlation coefficients were calculated using normalized data between 24 pI ranges which were aligned separately for each pair of samples and the Pearson correlation coefficients were calculated between the resulting aligned values. These were averaged for all distinct sample pairs within each of these three groups to produce a single within-group correlation coefficient for each group. Similarly, for every pair among these groups, correlation coefficients for every pair of samples spanning the two groups were averaged to produce a single between-group average correlation coefficient. These within-group and between-group correlation coefficients were prepared separately for each fraction, and also averaged across the fractions to provide an overall summary. It is interesting that the actual correlation numbers are such that correlation between the breast and ovarian clusters is greater than that of the ovarian carcinoma and OSE clusters, which in turn is greater than that of the breast and OSE clusters as represented in the dendrograms. It would be expected that the breast cancer and OSE clusters would have the lowest correlation.

There is some information on the molecular relationship of these cell lines based upon gene expression profiles in prior work (4, 30). There are some similarities in terms of the cell lines that cluster together compared to the protein mapping, but also very distinct differences. ISOE-144, IOSE-80 and HOSE-A cluster together by gene expression, but IOSE-80 distinctly does not cluster with these lines in the protein maps. Based on the mRNA expression data, ES-2 clear cell line clusters with the OSE samples while MDAH-2774 does not cluster with ES-2 or TOV-112D, but rather with the serous carcinoma cell lines. TOV-112D appears to cluster by itself in the gene expression arrays, whereas in the protein maps it clearly clusters with MDAH-2774 and ES-2. HEY and ES-2 were found to cluster in the gene expression data (4, 30), but HEY clustered with PEO1 in the protein expression data which makes sense since these are both serous carcinomas. OVCA429 and OVCA433 are closely clustered by gene expression as in the case of the protein mapping and OVCA432 is reasonably closely linked in both cases. OVCA420 is not closely linked to the OSE cell lines as in the protein expression maps. Nevertheless, it is not surprising that the gene expression and protein expression clusters provide different information since in prior work (28) it was found that the gene expression and protein expression for several cell lines had a poor correlation. Such poor correlation has been observed in other studies (31) and is due to the fact that many of the mRNA messages do not translate into proteins (32) or the proteins produced are short-lived or misfolded and rapidly degraded (33). Ultimately, though, it is the protein expression that determines the function of the cell so that the relationships as determined by protein mapping will be essential for searching for distinctive markers of cancer.

An important capability of the 2-D mapping technique is the use of proteomic patterns for classification using common marker bands in the comparison of different clusters of cell lines. The peak retention time and intensity for each band can be obtained using the Beckman software. As an example the average number of peaks for 11 samples for fraction pI 7.57-7.72 is 67. Based on the protein expression map, proteins can be classified into three groups. One group of proteins is likely to be common to most cell types. For this fraction, we found 12 proteins of all cells have the same retention time and are likely to be the same. A second set of proteins appears to be linked to one group of cell lines only. This set may provide the basis for detection and classification of serous carcinoma and have the potential to provide identifying biomarkers. A third group of proteins appears to be expressed uniquely on each individual cell line. It is possible to hypothesize that this third group of proteins is responsible for unique aspects of cell behavior.

In order to identify markers of groups of serous carcinoma cell lines, standardization and alignment of bands were performed and then comparisons were made separately at each hydrophobicity level between two groups of samples. A differentially expressed band is selected on the basis of having at least 4-fold different mean level within the two groups of samples. In addition, at least 25 consecutive hydrophobicity levels were required to meet these conditions in order for the band to be considered as a marker. Figure 7 shows peak patterns for two cluster samples, one group as OVCA429, OVCA433, and DOV13 and the other group as IOSE-80, PEO1, and HEY. The image is displayed in a format with each different sample on the x-axis and hydrophobicity on the y-axis. The relative intensities of the band are quantitatively proportional to the amount

of corresponding protein detected by UV absorption. The three groups of bands are only observed in the group IOSE-80, PEO1, and HEY but not in the group of OVCA433, OVCA429, and DOV13. After the identification with MALDI-TOF mass spectrometry combined with ESI-TOF-MS, we got the protein IDs are Dickkopf related protein-1 precursor, Histone H_2B and NACHT, LRR and PYD-containing protein 10. Figure 8 shows the total ion chromatogram of LCT with three small windows showing three spectrum of MALDI for each peak which corresponds to a single spot in Figure 7. Table 2 shows the detail information about identify these three proteins. The use of differential analysis allows us to identify proteins that may be common bands for classification and limits the number of proteins that might require further analysis by mass spectrometric techniques.

2.4 Conclusion

The use of the 2-D liquid mapping serves as a powerful tool for comparing the protein expression profiles of large numbers of samples. It provides an automated method for reproducibly running samples for interlysate comparisons. This strategy provides a means for comparing the profiles of different samples and classifying them according to the protein bands observed in their expression maps. This method has been used to classify a number of serous carcinoma cell lines compared to OSE lines and also several ovarian clear cell and endometrioid and breast epithelial cell lines. The results show that in most cases the serous carcinoma clustered within several groups while the breast cancer and other ovarian cell lines clustered separately. The method provides a means to search for markers that may serve to classify a specific set of cancers. It also reduces the potentially large amounts of protein expression data from a large number of samples into

a manageable data set. It thus reduces the number of significant bands that need to be identified by mass spec or other methods.

Cell Line	Description
HOSE-A	HPV16 E6/E7 immortalized ovarian surface epithelium (OSE)
IOSE-144	Life-extended (with SV-40 large T antigen) OSE cell lines
IOSE-80	Life-extended (with SV-40 large T antigen) OSE cell lines
HOC1	Ovarian serous carcinoma
PEO1	Ovarian serous carcinoma
DOV13	Ovarian serous carcinoma
HEY	Ovarian serous carcinoma
OVCA420	Ovarian serous carcinoma
OVCA429	Ovarian serous carcinoma
OVCA432	Ovarian serous carcinoma
OVCA433	Ovarian serous carcinoma
ES-2	Ovarian clear cell adenocarcinoma
MDAH-2774	Ovarian endometrioid adenocarcinoma
TOV-112D	Ovarian endometrioid adenocarcinoma
CAld	Fully malignant human breast cancer cells
Ca1a	Fully malignant human breast cancer cells
AT1	Preneoplastic human breast cells

Table 2.1. Cell lines used for comparison of ovarian cancer proteins

	protein	Accession #	MW(Da)/PI	MOWSE score	Mean Err ppm	% Cov	LCT-MW
A	Dickkopf related protein-1 precursor	O94907	28672/8.8	1.51E+03	-8.5	52	27920
B	Histone H2B	P14001	13896/10.3	9.72E+02	-6.7	30	13780
C	NACHT, LRR and PYD-containing protein 10	Q86W26	75033/6.8	8.52E+02	23	15	71560

Table 2.2 Proteins identified for differentially expressed spots.

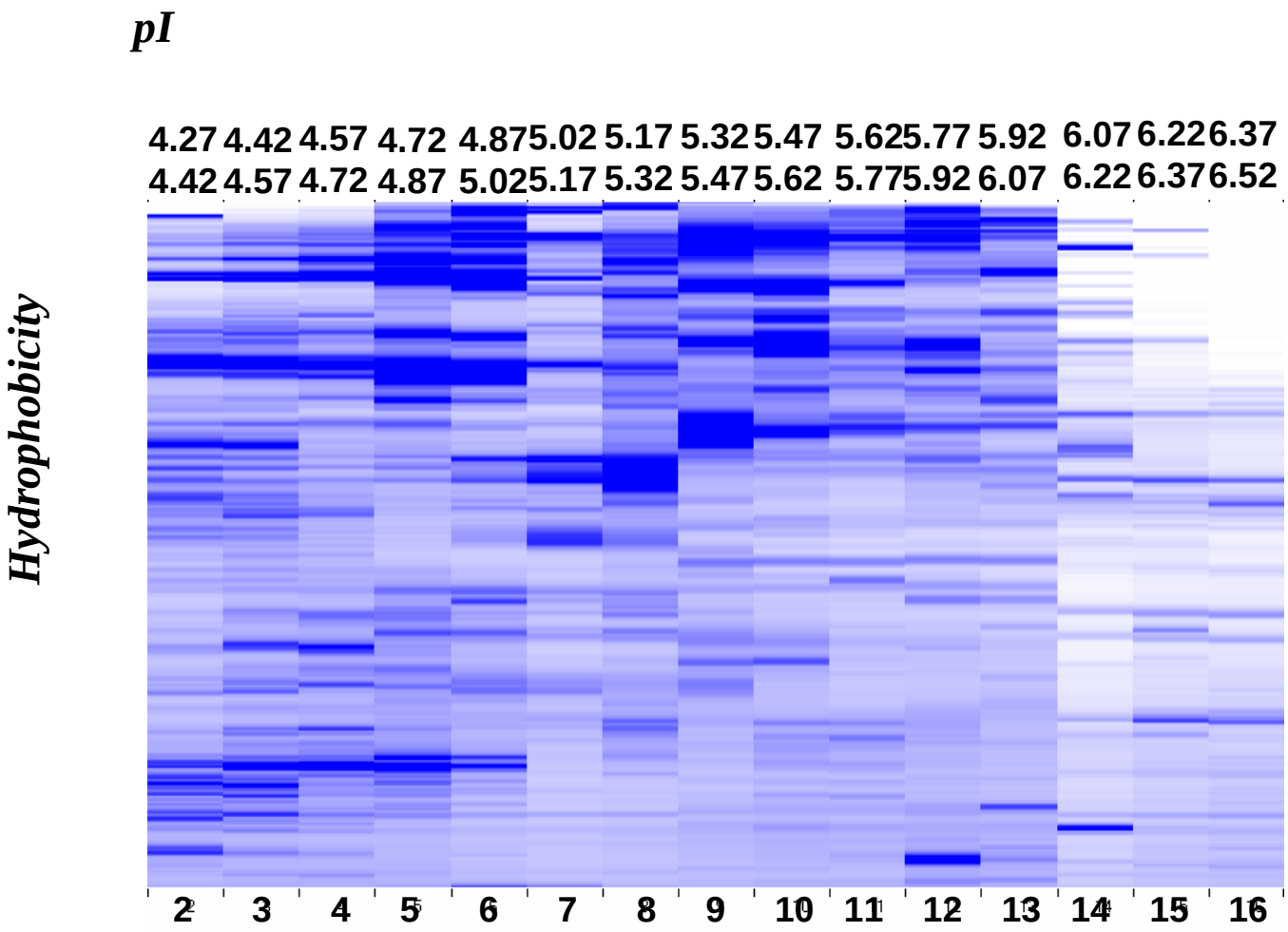


Figure 2.1 A blow-up of a portion of the 2-D liquid protein expression map of the ovarian cancer serous cell line HOC-1 whole cell lysate. Ovarian cancer serous cell line HOC-1 was separated using chromatofocusing over a pH range of 4.11 – 8.32 with 0.15 pH interval followed by separation in the second dimension using NPS RP-HPLC. The x-axis is pI of the chromatofocusing and the y-axis is hydrophobicity of the RP-HPLC. The scale of the bands represents the relative intensity of each band by UV detection at 214nm.

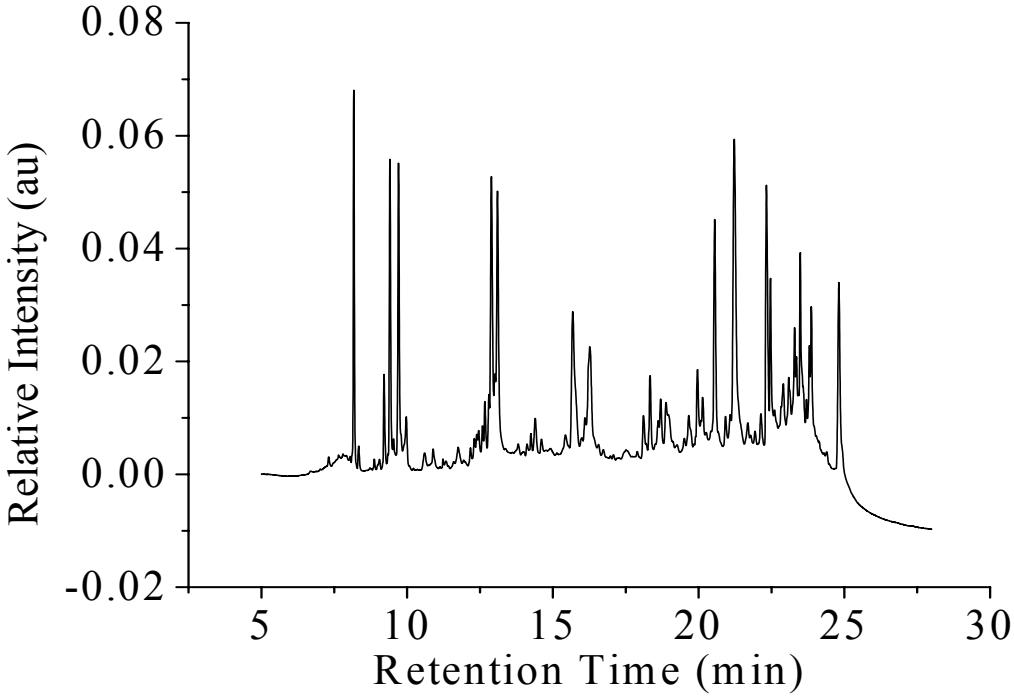


Figure 2.2 NPS RP-HPLC separation of one CF fraction Ovarian cancer serous cell line DOV13 pI fraction 4.85 -5.00 was separated by NPS RP-HPLC.

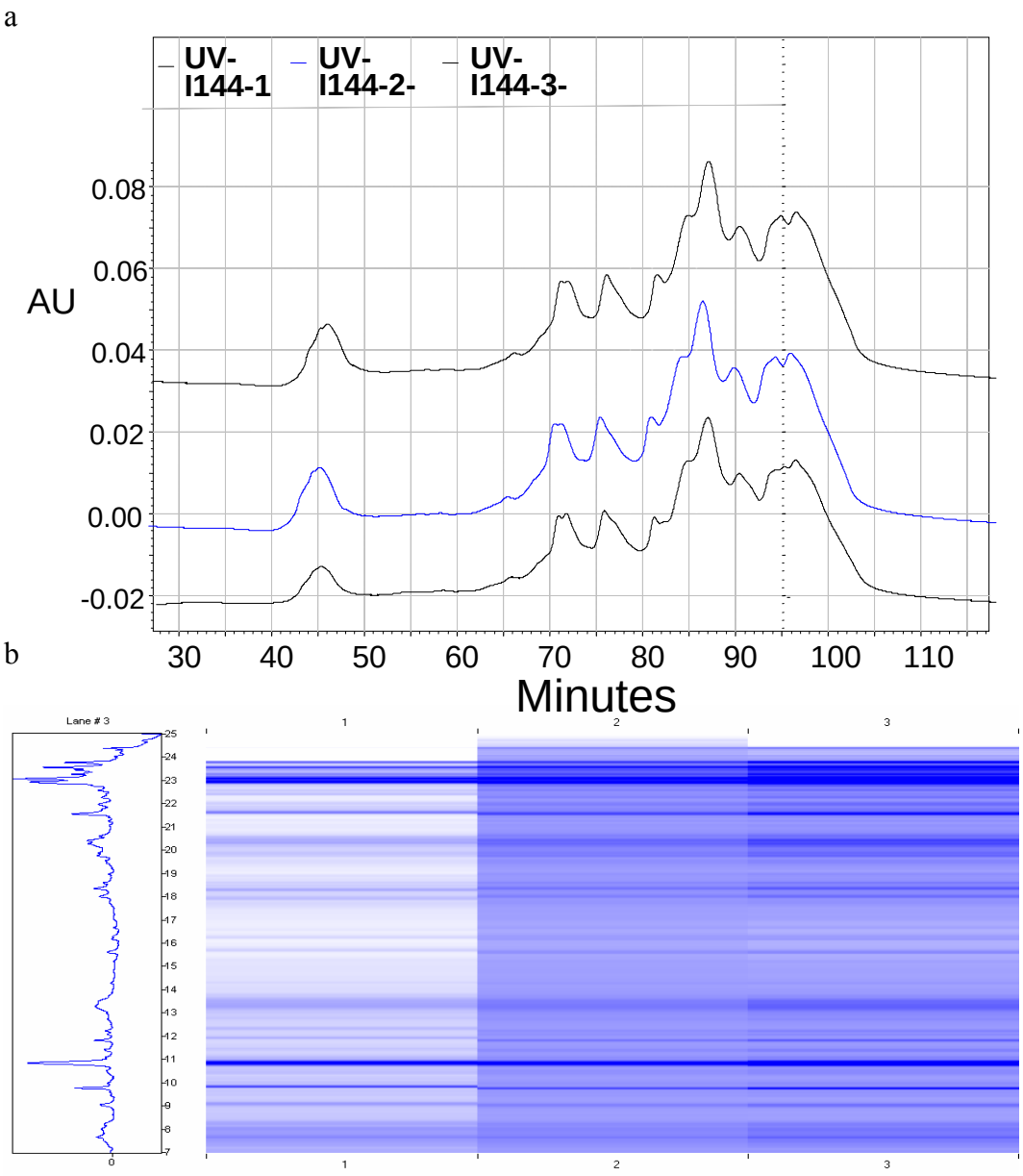
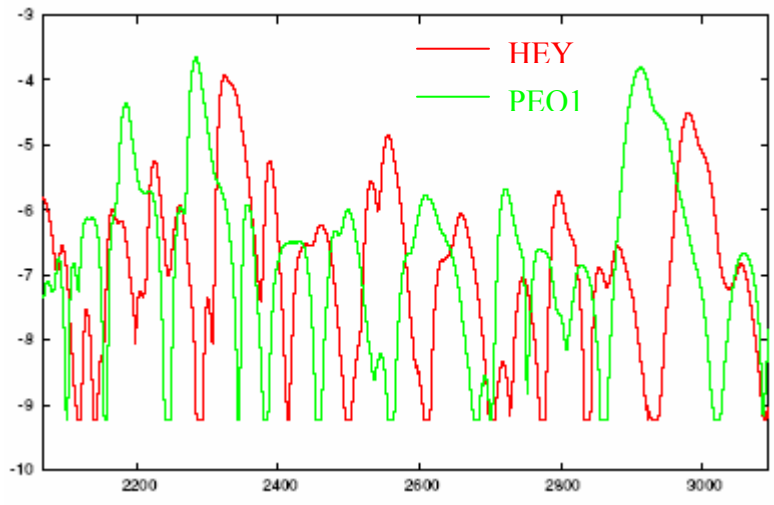


Figure 2.3 Reproducibility study of 2-D liquid separation. A. Chromatofocusing has been run three separate times on IOSE-144 sample, totally amount of sample was 3.5mg, 4.0mg and 4.5mg, where the protein content of the IOSE-144 cell lines is detected by UV absorption at 280nm. B. 2nd dimension has been run continuously by NPS RP-HPLC for each CF separation. One lane has been shown. UV absorption is 214nm.

A



B

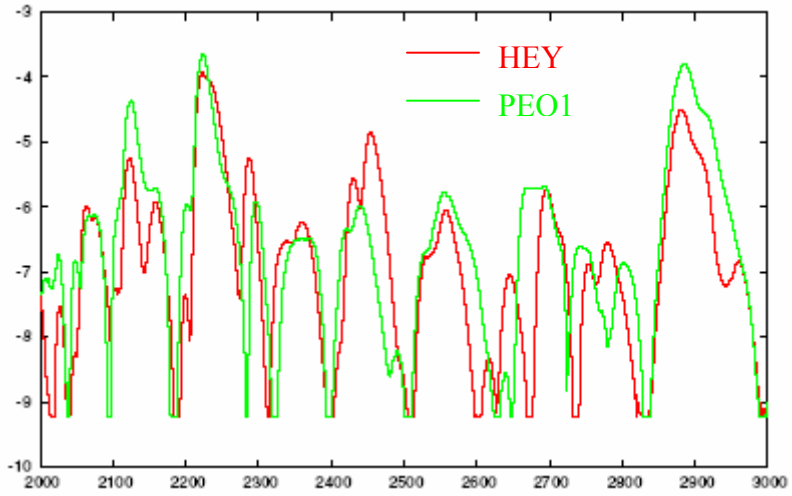
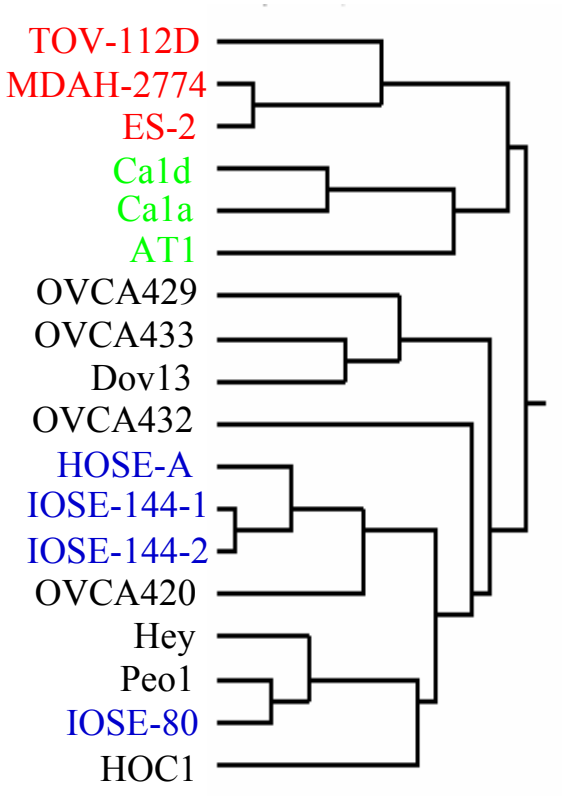
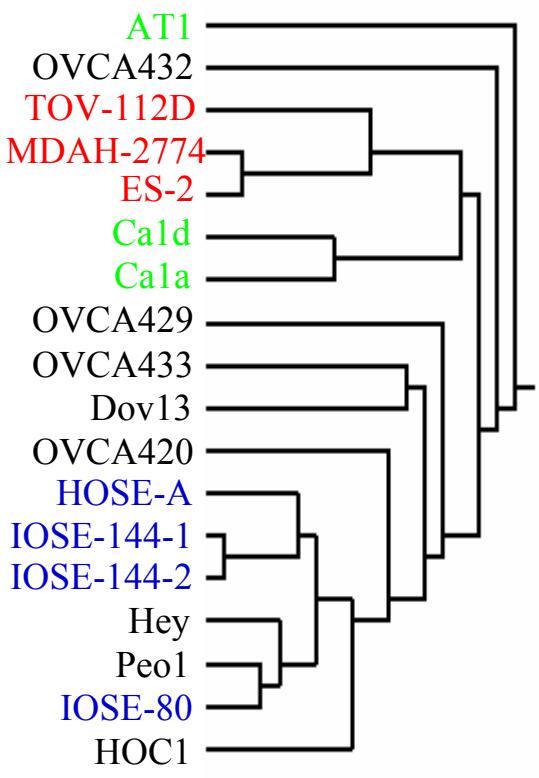


Figure 2.4. Alignment analysis. Superimposed sections of the chromatographic trace at 214nm from NPS RP-HPLC analysis of two serous samples, HEY and PEO1 fraction pI 4.55-4.70, gradient range from 41.5% to 49.5%. A.Prior to aligning B. After stretching of the entire profile

(A)


TOV-112D
MDAH-2774
ES-2
Ca1d
Ca1a
AT1
OVCA429
OVCA433
Dov13
OVCA432
HOSE-A
IOSE-144-1
IOSE-144-2
OVCA420
Hey
Peo1
IOSE-80
HOC1

(B)


AT1
OVCA432
TOV-112D
MDAH-2774
ES-2
Ca1d
Ca1a
OVCA429
OVCA433
Dov13
OVCA420
HOSE-A
IOSE-144-1
IOSE-144-2
Hey
Peo1
IOSE-80
HOC1

(C)

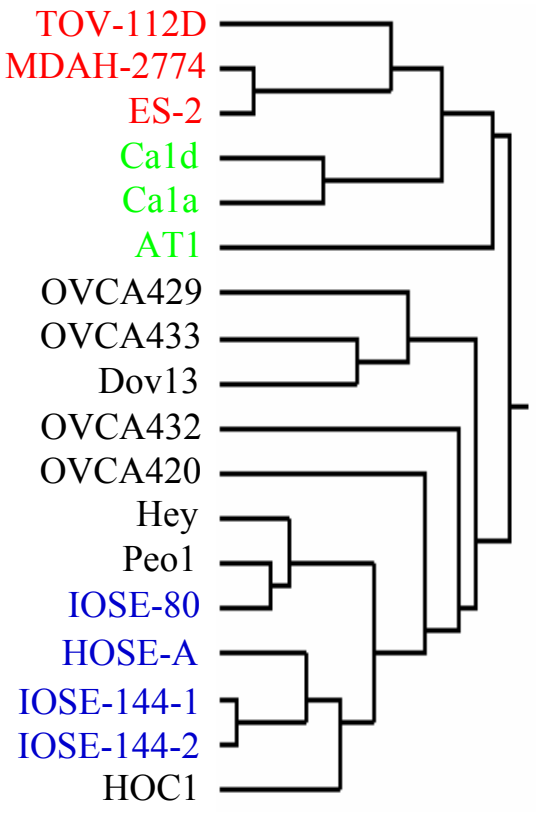


Figure 2.5 Hierarchical clustering analysis. Cell lines are aligned separately and grouped based on similarities in their protein expression using a hierarchical clustering analysis technique. This technique produces a dendrogram in which pairs of points are joined sooner (i.e. closer to the ends of the dendrogram) if they have greater correlation. A. Complete linkage B. Single Linkage C. Average Linkage.

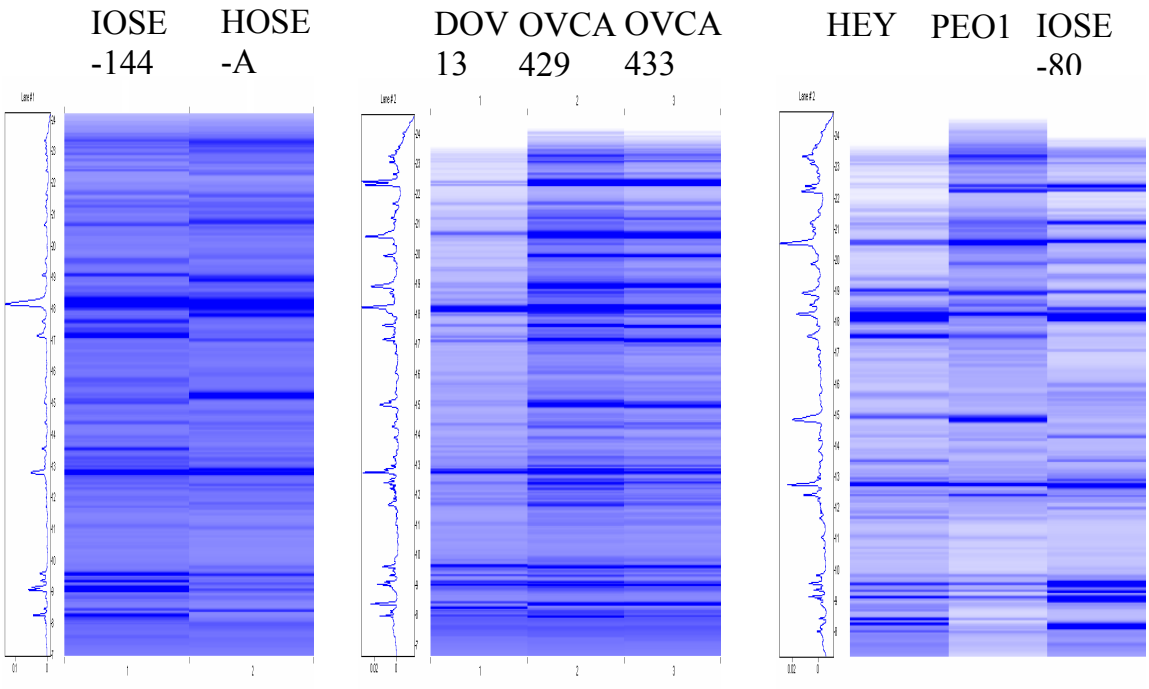


Figure 2.6 Protein expression maps for a given pH fraction. Fraction pI 4.55-4.70, gradient ranges from 30.0% to 78.0%. The relative intensities of the bands are quantitatively proportional to the amount of detected by UV absorption.

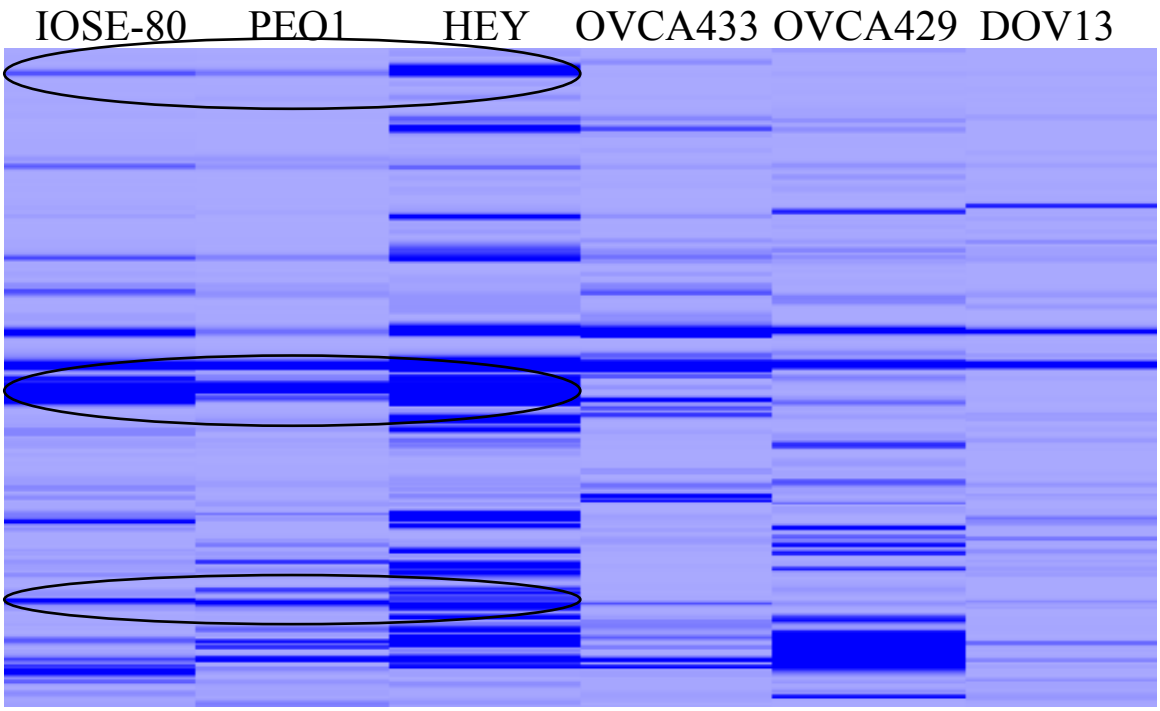


Figure 2.7 Differential analysis. Fraction pI 7.57-7.72, gradient ranges from 30.8% to 67.0%. The relative intensities of the bands are quantitatively proportional to the amount of protein detected by UV absorption.

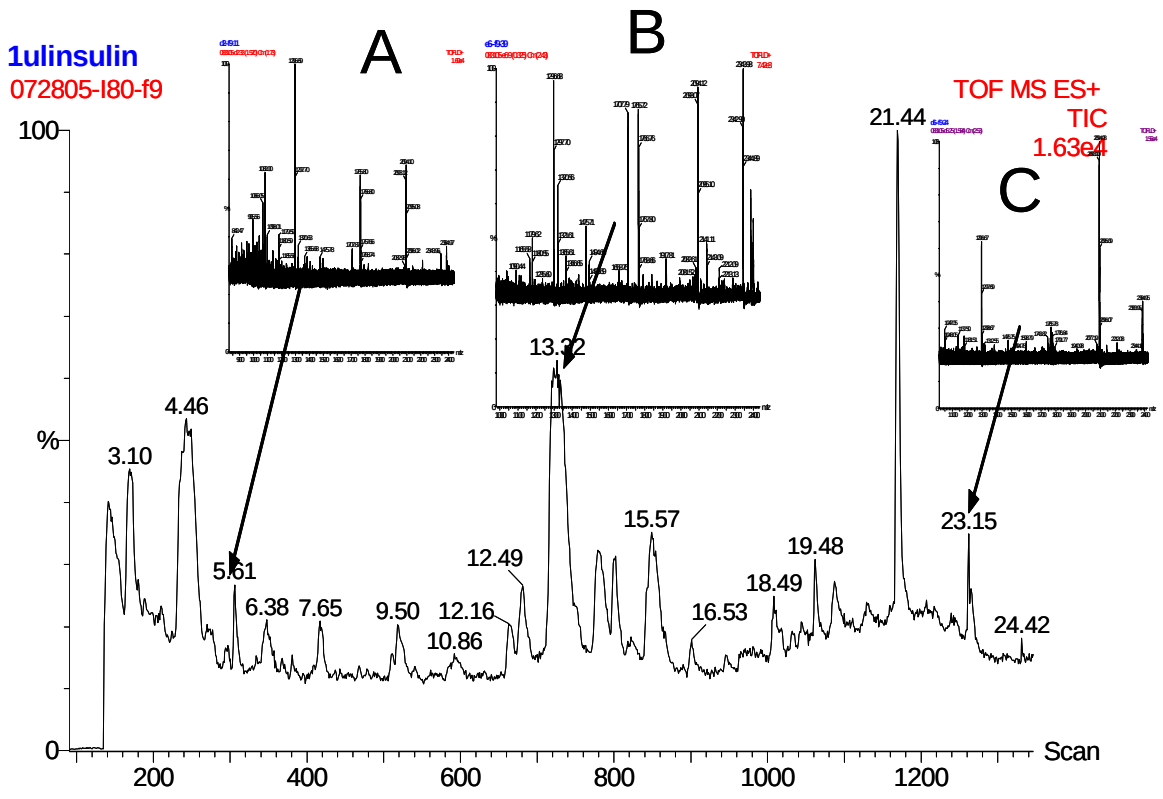


Figure 2.8 Three proteins identified for Figure 7.

2.5 References:

1. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537.
2. Alaiya, A. A., Franzen, B., Auer, G., and Linder, S. (2000) Cancer proteomics: From identification of novel markers to creation of artificial learning models for tumor classification. *Electrophoresis* **21**, 1210-1217.
3. Giordano, T. J., Shedden, K. A., Schwartz, D. R., Kuick, R., Taylor, J. M. G., Lee, N., Misek, D. E., Greenson, J. K., Kardia, S. L. R., Beer, D. G., Rennert, G., Cho, K. R., Gruber, S. B., Fearon, E. R., and Hanash, S. (2001) Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.* **159**, 1231-1238.
4. Schaner, M. E., Ross, D. T., Ciaravino, G., Sorlie, T., Troyanskaya, O., Diehn, M., Wang, Y. C., Duran, G. E., Sikic, T. L., Caldeira, S., Skomedal, H., Tu, I. P., Hernandez-Boussard, T., Johnson, S. W., O'Dwyer, P. J., Fero, M. J., Kristensen, G. B., Borresen-Dale, A. L., Hastie, T., Tibshirani, R., van de Rijn, M., Teng, N. N., Longacre, T. A., Botstein, D., Brown, P. O., and Sikic, B. I. (2003) Gene expression patterns in ovarian carcinomas. *Mol. Biol. Cell* **14**, 4376-4386.
5. Schwartz, D. R., Kardia, S. L. R., Shedden, K. A., Kuick, R., Michailidis, G., Taylor, J. M. G., Misek, D. E., Wu, R., Zhai, Y. L., Darrah, D. M., Reed, H., Ellenson, L. H., Giordano, T. J., Fearon, E. R., Hanash, S. M., and Cho, K. R. (2002) Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. *Cancer Res.* **62**, 4722-4729.
6. Alaiya, A. A., Franzen, B., Hagman, A., Dysvik, B., Roblick, U. J., Becker, S., Moberger, B., Auer, G., and Linder, S. (2002) Molecular classification of borderline ovarian tumors using hierarchical cluster analysis of protein expression profiles. *Int. J. Cancer* **98**, 895-899.
7. Jones, M. B., Krutzsch, H., Shu, H. J., Zhao, Y. M., Liotta, L. A., Kohn, E. C., and Petricoin, E. F. (2002) Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. *Proteomics* **2**, 76-84.
8. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Aksien, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000) Molecular portraits of human breast tumours. *Nature* **406**, 747-752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lonning, P. E., and Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10869-10874.
10. Korshunov, A., Neben, K., Wrobel, G., Tews, B., Benner, A., Hahn, M., Golanov, A., and Lichter, P. (2003) Gene expression patterns in ependymomas correlate with tumor location, grade, and patient age. *Am. J. Pathol.* **163**, 1721-1727.

11. Tan, Y. X., Shi, L. B., Tong, W. D., Hwang, G. T. G., and Wang, C. (2004) Multi-class tumor classification by discriminant partial least squares using microarray gene expression data and assessment of classification models. *Comput. Biol. Chem.* **28**, 235-244.
12. Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C. E., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., and Brown, P. O. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24**, 227-235.
13. Yanagisawa, K., Shyr, Y., Xu, B. G. J., Massion, P. P., Larsen, P. H., White, B. C., Roberts, J. R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J. H., Caprioli, R. M., and Carbone, D. P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* **362**, 433-439.
14. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-1053.
15. Alaiya, A. A., Franzen, B., Fujioka, K., Moberger, B., Schedvins, K., Silversvard, C., Linder, S., and Auer, G. (1997) Phenotypic analysis of ovarian carcinoma: Polypeptide expression in benign, borderline and malignant tumors. *Int. J. Cancer* **73**, 678-683.
16. Zhu, K., Zhao, J., Lubman, D. M., Miller, F. R., and Barder, T. J. (2005) Protein pI Shifts due to Posttranslational Modifications in the Separation and Characterization of Proteins. *Anal. Chem.* **77**, 2745-2755.
17. Kachman, M. T., Wang, H. X., Schwartz, D. R., Cho, K. R., and Lubman, D. M. (2002) A 2-D liquid separations/mass mapping method for interlysate comparison of ovarian cancers. *Anal. Chem.* **74**, 1779-1791.
18. Zhu, K., Kim, J., Yoo, C., Miller, F. R., and Lubman, D. M. (2003) High sequence coverage of proteins isolated from liquid separations of breast cancer cells using capillary electrophoresis-time-of-flight MS and MALDI-TOF MS mapping. *Anal. Chem.* **75**, 6209-6217.
19. Buick, R. N., Pullano, R., and Trent, J. M. (1985) Comparative Properties of 5 Human Ovarian Adenocarcinoma Cell-Lines. *Cancer Res.* **45**, 3668-3676.
20. Langdon, S. P., Lawrie, S. S., Hay, F. G., Hawkes, M. M., McDonald, A., Hayward, I. P., Schol, D. J., Hilgers, J., Leonard, R. C. F., and Smyth, J. F. (1988) Characterization and Properties of 9 Human Ovarian Adenocarcinoma Cell-Lines. *Cancer Res.* **48**, 6166-6172.
21. Kim, J. H., Skates, S. J., Uede, T., Wong, K. K., Schorge, J. O., Feltmate, C. M., Berkowitz, R. S., Cramer, D. W., and Mok, S. C. (2002) Osteopontin as a potential diagnostic biomarker for ovarian cancer. *J. Am. Med. Assoc.* **287**, 1671-1679.
22. ATCC Cell Biology Catalog.
23. Freedman, R. S., Pihl, E., Kusyk, C., Gallager, H. S., and Rutledge, F. (1978) Characterization of an Ovarian Carcinoma Cell Line. *Cancer* **42**, 2352-2359.
24. Provencher, D. M., Lounis, H., Champoux, L., Tetrault, M., Manderson, E. N., Wang, J. C., Eydoux, P., Savoie, R., Tonin, P. N., and Mes-Masson, A. M. (2000)

Characterization of four novel epithelial ovarian cancer cell lines. *In **Vitro** Cell Dev Biol Anim.* **36**, 357-361.
25. Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F., and Brooks, S. C. (1990) Isolation and Characterization of a Spontaneously Immortalized Human Breast Epithelial-Cell Line, Mcf-10. *Cancer Res.* **50**, 6075-6086.
26. Santner, S. J., Dawson, P. J., Tait, L., Soule, H. D., Eliason, J., Mohamed, A. N., Wolman, S. R., Heppner, G. H., and Miller, F. R. (2001) Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res. Treatment* **65**, 101-110.
27. http://www.ucl.ac.uk/oncology/MicroCore/HTML_resource/Hier_Linkage.htm.
28. Wang, H. X., Kachman, M. T., Schwartz, D. R., Cho, K. R., and Lubman, D. M. (2004) Comprehensive proteome analysis of ovarian cancers using liquid phase separation, mass mapping and tandem mass spectrometry: A strategy for identification of candidate cancer biomarkers. *Proteomics* **4**, 2476-2495.
29. Unpublished results Lubman Laboratory (2004).
30. Unpublished results Cho Laboratory, (2004).
31. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720-1730.
32. Woodman, A. C., Sugiyama, M., Yoshida, K., Sugino, T., Borgya, A., Goodison, S., Matsumura, Y., and Tarin, D. (1996) Analysis of anomalous CD44 gene expression in human breast, bladder, and colon cancer and correlation of observed mRNA and protein isoforms. *Am. J. Pathol.* **149**, 1519-1530.
33. Yewdell, J. W., Anton, L. C., and Bennink, J. R. (1996) Defective ribosomal products (DRiPs) - A major source of antigenic peptides for MHC class I molecules? *J. Immunol.* **157**, 1823-1826.

Chapter 3

Mass Mapping Study for Ovarian Serous Carcinoma Cell Lines

3.1 Introduction

Ovarian cancer is the most lethal gynecologic tumor in the United States and in the Western world. Based on morphological criteria, there are four major histological types of primary ovarian carcinomas (serous, clear cell, endometrioid, and mucinous).Among all these histological subtypes of ovarian cancer, serous carcinoma is the most common one and accounting for nearly 50% of all ovarian epithelial tumors (1). Most ovarian epithelial tumors are believed to be derived from the ovarian surface epithelium (OSE) (2, 3) and consist of benign, low malignant epithelial tumors, and invasive carcinomas (4, 5). Although surgery followed by chemotherapy is able to induce clinical responses in the majority of ovarian cancer patients, most of these women eventually die from the development of chemotherapy-resistant disease (1). It is very important for the identification of novel molecular markers which is useful for the early diagnosis and therapy of ovarian cancer patients. A better understanding of the molecular expression in normal ovarian samples and in ovarian cancer will likely provide promising targets for novel therapeutic strategies.

In order to identify markers with a differential pattern of expression between ovarian tumors and normal ovarian cells and to use this knowledge for the development of novel diagnostic and therapeutic markers against this disease, many studies have been

applied to compare the expression profiles of ovarian cancer to those of normal ovaries (6-9). Although cancer can be described as a disease of genes, mRNA and protein expression from a given gene may be discordant and it is the protein products of genes that actually determines the function and structure of the cells. The use of protein expression may be more informative and could lead to biologic insights and potential improvements in diagnosis and treatments for this disease. LC-MS based multi-dimensional separation methods are rapidly emerging. This technique can apply large scale protein expression profiling analysis and identify the relative quantities of the various proteins between different samples (10-13). Our group has developed a 2-D liquid phase mass mapping method which can explore protein-based expression profiling for ovarian cancer classification and to search for subtype-specific biomarker candidates (14). The method involved fractionation according to p*I* using chromatofocusing in the first dimension followed by separation of the proteins in each p*I* fraction using nonporous RP HPLC, which was coupled to an ESI-TOF mass analyzer for molecular weight (MW) analysis. A 2-D mass map of the protein content of each ovarian cancer samples based upon p*I* versus intact protein MW was generated.

In the present work, we demonstrate the capabilities of using the 2-D liquid phase separation of protein method combined with on-line ESI-TOF mass spectrometry for mapping five ovarian serous carcinoma cell lines and one epithelial cell line IOSE-80. It is shown that using this method, we have mapped several interesting pI fractions and have plotted them as a 2-D map of pI versus accurate MW. Some interesting proteins have been collected in the liquid phase and analyzed by MALDI-TOF mass spectrometry, where their confident ID can be obtained by combing both intact molecular weight and

peptide fingerprint. The comparison can be made quantitative by the use of internal standards to normalize for the changes in ESI intensity. Clustering can be applied to those cell lines based on their mass map and a comparison can be made between different groups based on the clustering results.

3.2 Experimental Section

3.2.1 Cell Culture and Sample Preparation

A total of six ovarian cell lines were applied in this study, which included five serous and one immortalized ovarian surface epithelial cell line. Ovarian serous carcinoma-derived cell lines HEY were a gift from L. Dubeau (USC School of Medicine, Los Angeles, CA); DOV13, OVCA429 and OVCA433 were a gift of D. Fishman (North-western University, Chicago, IL); PEO1 was a gift from T. Hamilton (Fox Chase Cancer center, Philadelphia, PA). Ovarian surface epithelial cells expressing SV40 large T Antigen (IOSE-80) were a gift of N.Auersperg (University of British Columbia, Vancouver, British Columbia, Canada). Cells were cultured as previously described (15). All the cell lines were cultured under standard conditions in the Department of Pathology at the University of Michigan Medical School. The cells were maintained in DMEM/10% fetal bovine serum and 1% Penicillin/Streptomycin (Invitrogen, Gaithersburg, MD). When monolayer cultures were 80-90% confluent, the cells were washed three times with phosphate buffered saline (PBS) and lysed in lysis buffer. Lysis buffer (90 $\mu L/cm^2$) consisted of 6 M urea (INC Biochemicals, Cleveland, OH, USA), 2M thiourea (INC), 1% n-octyl-β-D-glucopyranoside (OG1) (Sigma, St. Louis, MO, USA), 2mM DTT (Sigma), 2% Biolyte ampholytes pH 3-10 (Bio-Rad, Richmond, CA, USA), and 2.5 mM PMSF (Bio-Rad). The cells were removed by scraping with a cell scraper (Costar, Cambridge,

MA, USA). The insoluble material was removed by ultracentrifugation at 35000 rpm at 4 ºC for 1h, and the supernatant was stored at -80ºC for further use.

3.2.2 Chromatofocusing(CF)

CF was performed using ProteomeLab™ PF 2D system (Beckman Coulter, Inc. Fullerton, CA, USA). An HPCF-1D column (250 × 2.1 mm) was used to perform chromatofocusing. Two buffers, a start buffer (SB) (Beckman Coulter, Inc. Fullerton, CA, USA) and an elution buffer (EB) (Beckman Coulter, Inc. Fullerton, CA, USA), were used to generate the pH gradient on the column. Both buffers were prepared in 6 M urea and 0.2% octyl glucoside. Before running the CF, the pH of SB was adjusted to 8.5 +/- 0.1 and EB was adjusted to 4.0+/- 0.1 using either a saturated solution (50 mg/mL) of iminodiacetic acid (Sigma, part # I5629) if the buffer was too basic or 1 M NH_4OH if the buffer was too acidic. A PD-10 G-25 column (Amersham Pharmcia Biotech) was used to exchange the protein sample from the lysis buffer to the equilibration buffer used in the CF experiment.

The HPCF 1D column was first flushed with 100% distilled water (filter through a 0.45 um filter) for 10 column volumes at 0.2 mL/min, then equilibrated with 100% SB for 30 column volumes. After equilibration with SB, the HPCF column was ready to start the ProteomeLab PF 2D default method where injection of the sample began the method. After the method had been started, the column was washed with 100% SB to remove material that did not bind to the column at pH 8.5. When the wash was complete, the UV absorbance returned to baseline. Once a stable baseline was achieved, the method was initiated at 100% EB. UV detection was performed at 280 nm and the pH was monitored on-line by a flow-through pH probe (Beckman Coulter, Inc. Fullerton, CA, USA). As the

pH decreased, pH fractions were then collected in 0.15 pH intervals where 30 fractions in total were collected in the range of pH 8.5 - 4.0. After the pH of the eluent reached 4.0, the HPCF column was washed with 10 column volumes of 1M NaCl and the fractions collected by time. After the salt wash, the HPCF column is washed with 10 column volumes of distilled or deionized water. The CF portion of the method for the ProteomeLab PF 2D required around 185 minutes.

3.2.3 LC/ESI-TOF-MS

When the first-dimension separation was completed, the pI fractions collected from the first dimension were then run on the second dimension. Proteins were resolved by reversed-phase chromatography using a HPCF-2D (4.6 × 33 mm) NPS column (Beckman Coulter, Inc. Fullerton, CA, USA). The LC separation was performed at a flow rate of 0.4 mL/min and the column temperature was 65 ºC.. Solvent A was 0.1% TFA and 0.3% formic acid (FA) in H_2O and solvent B was 0.08% TFA and 0.3% formic acid (FA) in. The Gradient was run from 15 to 25% B in 1 min, 25 to 35% in 6 min, 35 to 38% in 4 min, 38 to 45% in 6 min, 45 to 65% in 2 min, 65 to 67% in 6 min, and finally up to 100% in 1 min, then back to 5% in 1 min. The column was equilibrated with 5% of solvent B until the A of UV (214 nm) (Thermo separation products sp UV 200, FL) became stable and the protein fraction from CF separation was applied to the column. A postcolumn splitter was used to split the effluent into half. Half of the effluent was delivered into the source of Micromass LCT™ workstation (Waters, Milford, MA, USA). The other half of the effluent was passed through a Beckman model 166 UV absorption detector (214 nm) and collected using a Beckman SC-100 fraction collector for MALDI-TOF MS analysis. The LCT parameters were set as follows: desolvation temperature

300ºC, source temperature 110ºC, desolvation gas 970 L/h, capillary voltage 13200 V, sample cone 135 V, extraction cone 12 V, and rf voltage 750 V. Before sample injection, the external calibration was utilized with the direct infusion of NaI-CsI solution by a syringe pump. Bovine insulin (0.25 ug) was mixed with samples and injected into the LC-MS for calibration and quantification. Maxent1 software was used for LCT data analysis. The total ion chromatogram (TIC) was scanned for regions that contain redundant multiple-charged peaks, and those regions were combined for deconvolution. The deconvoluted peaks were subsequently combined into a single mass spectrum, which was converted to a text file for input into the 2-D mapping software (Proteovue) written in-house. Data is processed using Micromass MassLynx™ v3.4 software (Waters), and protein multiply-charged umbrellas are deconvoluted using Micromass MaxEnt® 1 software (Waters). Deconvolution is performed using a mass range of 5–85 kDa, 1 Da resolution, 0.75 Da peak widths, and a 65% peak height value.

3.2.4 Tryptic digestion of the LCT fractions

Fractions collected from the LCT separation were concentrated down to 20 mL using a Speedvac concentrator (Labconco, Kansas City, MO). To the reduced volume fractions,10% v/v 10 mM DTT (Sigma), 10 mL 1M NH_4HCO_3(Sigma) and 0.25 mg of TPCK-treated trypsin (Promega, Madison, WI) were added. The fractions were then incubated at 37ºC for 24 h. After 24-h incubation, 1 ul TFA was added into each fraction to stop digestion.

3.2.5 MALDI-TOF-MS Sample Preparation and Data Acquisition

Before MALDI analysis, each tryptic digested sample was desalted and concentrated using C18 ZipTip (Millipore) into 5 μL 60% (v/v) ACN/0.1% (v/v) TFA.

One microliter of the concentrated peptide mixture was spotted onto a Micromass 96-spot plate followed by 1 μL of matrix-standard layered on top of it. The MALDI matrix solution was prepared by diluting saturated α-CHCA (Sigma) solution with 60% v/v ACN and 0.1% v/v TFA at 1:4 ratio v/v with three internal standard proteins: angiotensin I (1296 Da), adrenocorticotropic hormone (ACTH clip 1–17, 2093 Da), and ACTH clip 18–39 (2465 Da) (Sigma). 1 mg/mL standard proteins were diluted 100-fold with deionized water. These standards were further diluted 50, 40, and 30 times, respectively, with the diluted matrix.

The Micromass TofSpec-2E™ (Waters) was used to generate peptide mass fingerprints (PMF), followed by database searching in the SwissProt database. Spectra were calibrated using the internal calibrants to achieve a mass accuracy of less than 50 ppm. Monoisotopic peak lists were generated and searched against the SwissProt protein database, using the MS-Fit search engine (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm). The search was performed using the following parameters: (1) species: human; (2) allowing one missed cleavage; (3) possible modifications: peptide *N*-terminal glutamine to pyroglutamic acid, oxidation of methionine, and protein *N*-terminus acetylated and phosphorylation of S, T and Y; (4) peptide mass tolerance 50 ppm; (5) MW ranged from 1000 to 100,000 Da; (6) p*I* range of protein 3–10. Protein identifications were considered confident matches when meeting the following criteria: (i) ranked in the top three database hits; (ii) sequence coverage is >20%; (iii) average mass error is <50 ppm (iv) MOWSE scores over 10^4.

3.2.6 Clustering analysis

Protein abundances for each sample were first log (base 2) transformed and summed with consecutive, non-overlapping windows of width 26104 PPM. That is, starting at a lower mass bound of 5000 Da, the first window spanned from 5000 to 5000(1 1 26104/106) Da = 5100 Da, the second window spanned from 5100 Da to 5100(1 1 26104/106) Da = 5202 Da, and so on. At a typical protein mass of 50 kDa, the mass window was 1000 Da wide, with wider windows at greater masses and narrower windows at lesser masses. Absolute measurement error is greater for larger proteins, and larger proteins are presumed to have greater mass variation due to PTMs. Based on the mass maps, clustering can be applied according to the clustering analysis using in-house software Pathon. Next, two-sample *t*-tests were performed within each window, to compare the mean expression level in one group from the mean expression level in the other group. A significant difference was defined when the *t*-test *p*value was less than 0.01 and when the levels within each group had standard deviation greater than 1026. For display purposes , bands with significantly greater expression in one group were colored pink and bands with significantly greater expression for another gorup were colored cyan. Bands in which a nonzero abundance was detected in every sample were colored yellow.

3.3 Results and Discussion

3.3.1 Cell lines

Five ovarian serous carcinoma-derived cell lines and one OSE cell lines IOSE-80 were used to study ovarian cancer proteomes. Their description was shown on Table 1.

3.3.2 Analysis of ovarian cancer proteomes

A 2-D liquid phase mass mapping method has been developed in our laboratory and applied to profile protein expression in ovarian serous cancer cell lines. Proteins extracted ovarian carcinoma samples were separated by utilizing this 2-D liquid phase separation method and detected intact molecular weight in order to do the classification and comparison based on their protein expression. This analytical approach fractionated proteins by CF in the first dimension based on *pI*, and each pH fraction was further separated by nonporous (NPS)-RP-HPLC in the second dimension. The protein eluents from NPS-RP-HPLC were interfaced online to ESI-TOF to obtain an accurate and reproducible protein intact molecular weight. The 2-D mass maps of each sample were finally generated by the Proteovue or Deltavue software, which also allows for highly accurate protein profile comparisons between the two different groups of ovarian cancer.

Figure 1 shows a representative chromatograph we obtained using the LCT as a detector. In each injection, we inject 0.25ug insulin as an internal standard. We can thus normalize the data to compensate the small variance due to different instrumental conditions. In this figure, we indicate a small peak with retention time 10.602. Compared to the insulin peak, it is quite small; however, after mass combination and deconvolution, we can still obtain a good molecular weight value and which shows the excellent sensitivity of this method.

Figure 2 shows the 2-D comparison mass map of 4 serous ovarian cancer samples generated from a selected pH fraction (5.15–5.30). The horizontal and vertical dimensions of the 2-D protein content map are plotted in terms of the selected *pI* range and protein intact MW, respectively. Each vertical lane on the map represents the protein

content of the same selected pH fraction collected from each sample. Each band represents a protein eluted from RP-HPLC and detected by ESI-TOF. From the mass expression, we observe that there are differences and similarities between these four samples. Some proteins are the same for all four samples and some are unique for each sample. As mentioned in the methods, a postcolumn splitter was used to split the effluent after HPLC separation. Half of the effluent was delivered into the source of the Micromass LCT™ workstation and the other half of the effluent was passed through a UV detector and collected for MALDI-TOF MS analysis. Combining the information from mass spectrometry, we can obtain a confident ID based on the intact molecular weight and peptide fingerprint. Table 2 shows the protein identified from Peo1 fraction pI 5.15-5.30. We also studied the similarities between these four serous cell lines. Table 3 shows the common proteins identified for these four samples. The LCT intact molecular weight is listed for each sample. Most of these proteins have the same intact molecular weight for four samples which means this method is very reproducible. The small variance may due to the small shift of the calibration. Figure 3 shows the comparisons of four serous carcinoma samples' intact protein molecular weight obtained by electrospray TOF-MS fraction pI 4.40-4.55, pI 4.85-5.00 and pI 5.00-5.15. These 2-D maps provide an overview of protein mass distribution and composition in each fraction.

An important capability of the 2-D mapping technique is the use of proteomic patterns for comparison to find those differentially expressed proteins which may serve as potential makers. In Figure 4, the IOSE80 cell line is displayed with the protein bands in shades of red, while the OVCA429 cell line is displayed with the protein bands in shades of green. The lane in the center is the differential display where the intensities of the

protein bands of the same MW have been subtracted from one another. There are some proteins which are highly expressed in the normal cell line while some proteins are highly expressed in cancer cell lines. After identification using both intact molecular weight and peptide fingerprint, we marked the differentially expressed proteins in the figure.

3.3.3 Hierarchical clustering analysis

The hierarchical clustering method was used for classification on six ovarian cell lines based on their similarities in protein expression. This technique produces a dendrogram in which pairs of points are joined sooner (*i.e.* closer to the ends of the dendrogram) if they have greater correlation. In-house software Pathon was used to do the classification. Three different cluster analyses: "single linkage," "complete linkage," and "average linkage" hierarchical clustering were used. Differences between these three methods arise because of the different ways of defining distance (or similarity) between clusters. Two different fractions were applied in Figure 5. From the results we observed that some specific cell lines are grouped together.

3.4 **Comparison of expression patterns distinguishes different groups**

The comparison between two different groups can be viewed using the mass mapping technique. As in Figure 6, group OVCA432, OVCA429 and DOV13 were compared with group HEY, PEO1 and IOSE80. For identification of differentially expressed proteins between the two different groups, the protein mass range was divided into windows of 20 000 PPM widths; that is, a window that starts at K Daltons ends at K 1 K/20 Daltons. This window was chosen to include possible isoforms due to posttranslational modifications. Other windows of 10 000 PPM and 5000 PPM were also tested and, generally, similar results were obtained(data not shown). A non-overlapping

sequence of windows was constructed covering the entire measured range. Within each of these windows, the total measured protein amount was summed for each sample, and these sums were compared between the samples using two-sample *t*-tests. Figure 6 shows the results of this analysis for two selected p*I* range. Pink bands in the figure correspond to mass ranges that are significantly more abundant in group OVCA432, OVCA429 and DOV13. Blue bands correspond to mass ranges that are significantly more abundant in group HEY, PEO1 and IOSE80. Yellow bands correspond to mass ranges in which all samples have some level of measured protein expression.

3.4 Concluding Remarks

A 2-D liquid phase mass mapping method has been applied to explore variations in protein expression related to ovarian serous carcinoma and normal OSE cell lines. The 2-D p*I*/mass maps were generated and used for interlysate comparison of protein expression between the different samples. Using this technique, we obtain pI, intact molecular weight and hydrophobicity of samples which provides us more information. This mass mapping methods can be used to classify ovarian serous carcinoma cell lines and the mass map can be used as a fingerprint for the proteome. The differential display map can show the presence of up- or down-regulated proteins, where some interesting differentially expressed proteins between IOSE80 and OVCA429 were identified by PMF as promising candidates of biomarkers. Identification of these markers should eventually facilitate future studies aimed at improving the diagnosis and treatment of ovarian cancers.

Cell line	Description
IOSE-80	Life-extended (with SV40 large Tantigen) OSE cell lines
PEO1	Ovarian serous carcinoma
HEY	Ovarian serous carcinoma
DOV13	Ovarian serous carcinoma
OVCA429	Ovarian serous carcinoma
OVCA433	Ovarian serous carcinoma

Table 3.1 Cell lines used for comparison of ovarian cancer proteins

Protein Name	MsFit pI	MsFit MW	LCT MW	Access #	MOWSE Score	% Cov
Proto-oncogene C-crk	5.5	33872	33852	P46108	4.47E+03	27
Tumor protein D54	5.3	22238	22287	O43399	1.14E+07	50
Keratin, type II cytoskeletal 7	5.5.	51418	51351	P08729	1.47E+07	37
Keratin, type I cytoskeletal 9	5.1	61988	-	P35527	2.15E+04	17
Keratin, type I cytoskeletal 19	5.0	44106	43998	P08727	2.54E+05	30
Serine/threonine protein kinase 4	5.0	55631	55447	Q13043	4.56E+04	20
UMP-CMP kinase	5.4	22223	22222	P30085	1.66E+04	33
Crk-like protein	6.3	33777	33800	P46109	2.94E+06	41
Keratin, type I cytoskeletal 18	5.3	48058	47966	P05783	5.14E+04	20
Keratin, type II cytoskeletal 8	5.5	53675	53621	P05787	1.99E+09	36
Heat shock cognate 71 kDa protein	5.4	70899	70877	P11142	5.76E+04	14
Moesin	6.1	67821	67700	P26038	1.21E+05	25
Lamin A/C	6.6	74140	-	P02545	1.33E+03	16
Stathmin	5.8	17303	17293	P16949	1.02E+03	63

Table 3.2 Protein identified from Peo1 fraction pI 5.15-5.30

Protein ID Fraction 5.15-5.30	MsFit pI	MsFit MW	LCT MW PEO1	LCT MW OVCA429	LCT MW OVCA433	LCT MW DOV13
Lamin A/C	6.6	74140	74766	74766	74766	-
Heat shock cognate 71 kDa protein	5.4	70899	70882	70882	70882	70882
Keratin, type II cytoskeletal 8	5.5	53675	53628	53628	53628	53628
Keratin, type II cytoskeletal 7	5.5	51418	51876	51876	51876	51876
Keratin, type I cytoskeletal 18	5.3	48058	47979	47979	47979	47979
Keratin, type I cytoskeletal 19	5.0	44106	44106	44106	44106	44106
Actin, cytoplasmic 2	5.3	41793	41722	41722	41722	41722
Crk-like protein	6.3	33777	33800	33800	33800	33800
UMP-CMP kinase	5.4	22223	22222	22222	22222	-
ATP synthase D chain, mitochondrial	5.2	18491	-	18400	18402	18402
Stathmin	5.8	17303	17293	17298	17293	17293
Astrocytic phosphoprotein PEA-15	4.9	15040	14978	14978	14978	14978
Alpha-endosulfine Alpha-endosulfine	6.6	13389	13380	13380	13380	13375
Histone H4	11.4	11367	11350	11350	11355	11355

Table 3.3 Common proteins identified for 4 samples

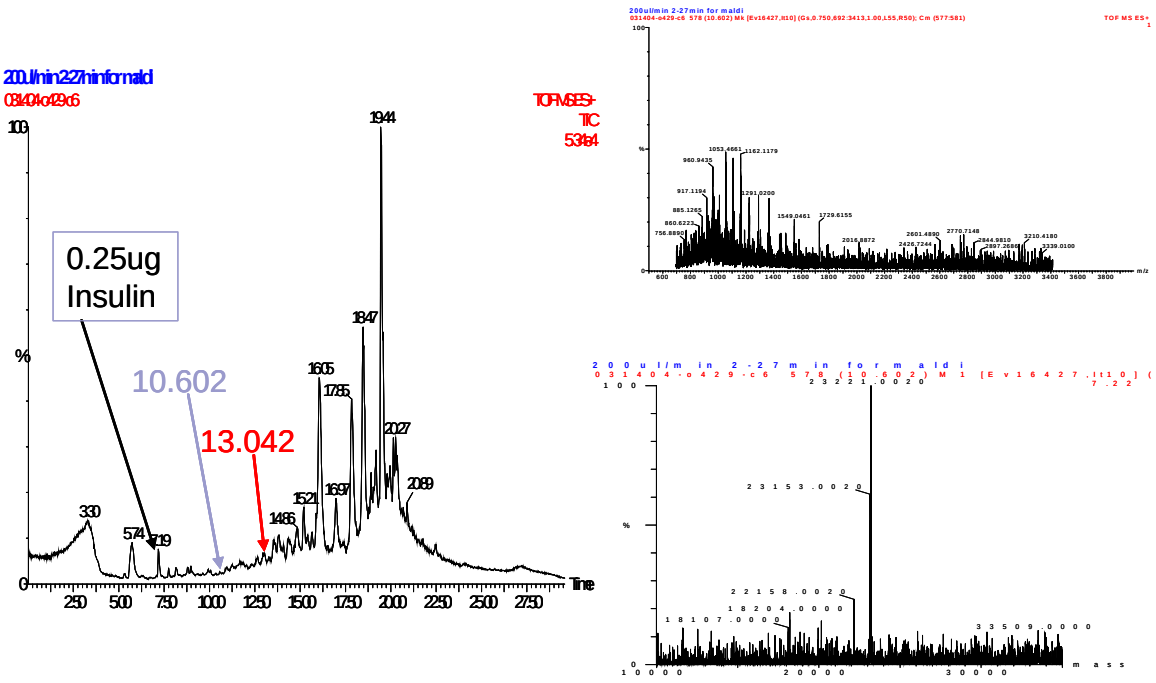


Figure 3.1 Online NPS RP HPLC-ESI TOF

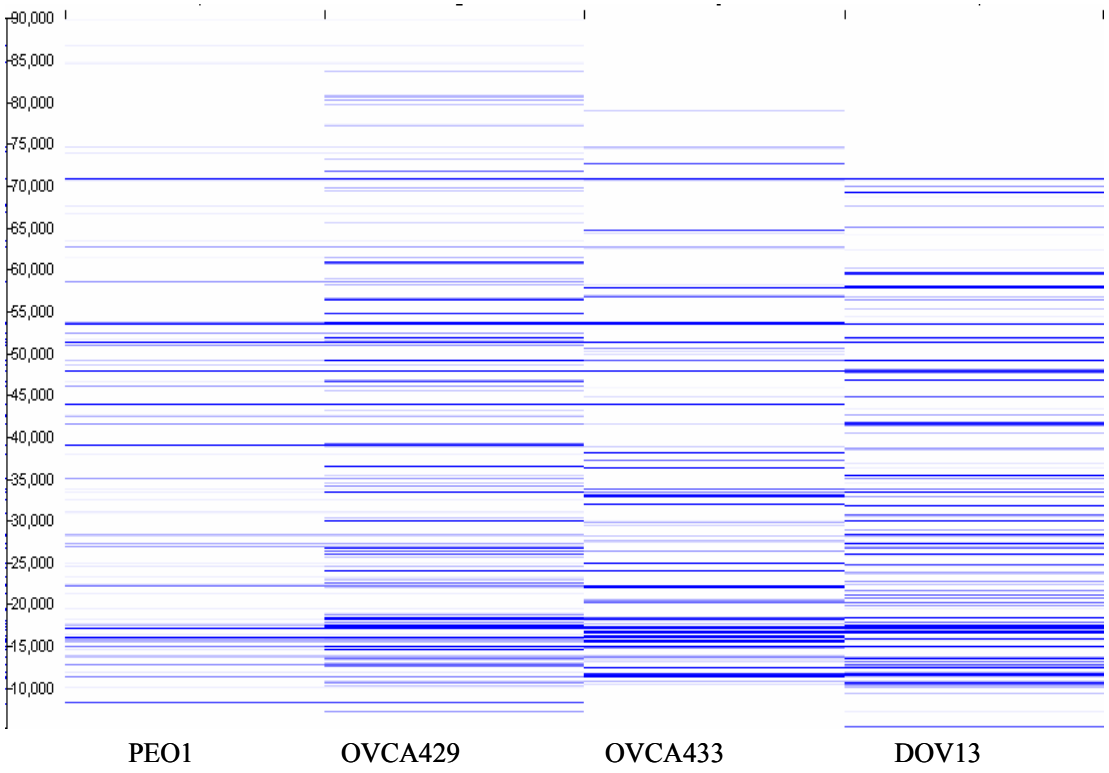


Figure 3.2 Mass map of four serous carcinoma: pI range 5.15-5.30

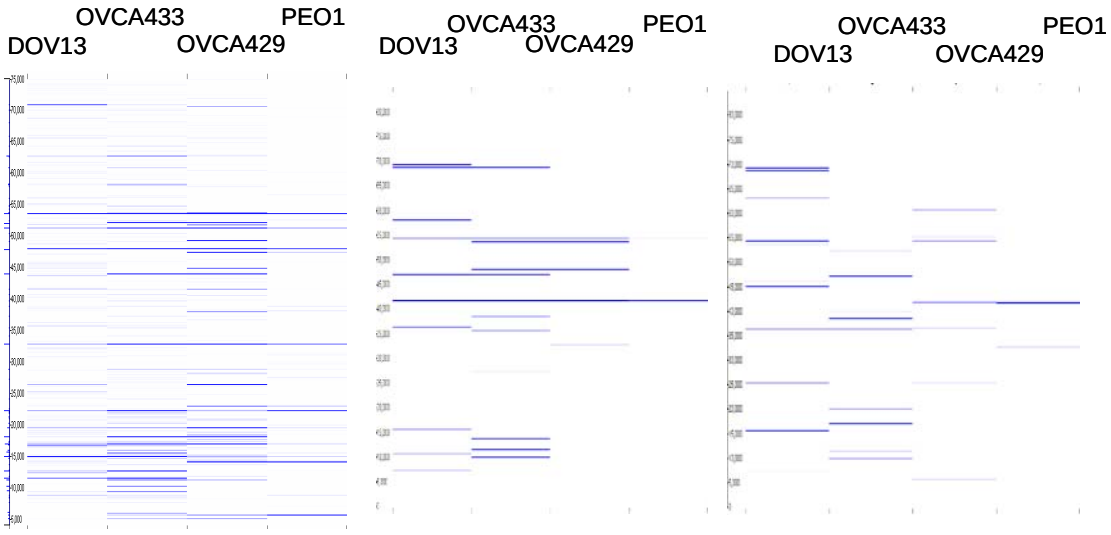


Figure 3.3 Comparisons of four serous carcinoma intact protein molecular weight obtained by electrospray TOF-MS pH 4.40-4.55 pH 4.85-5.00 pH 5.00-5.15

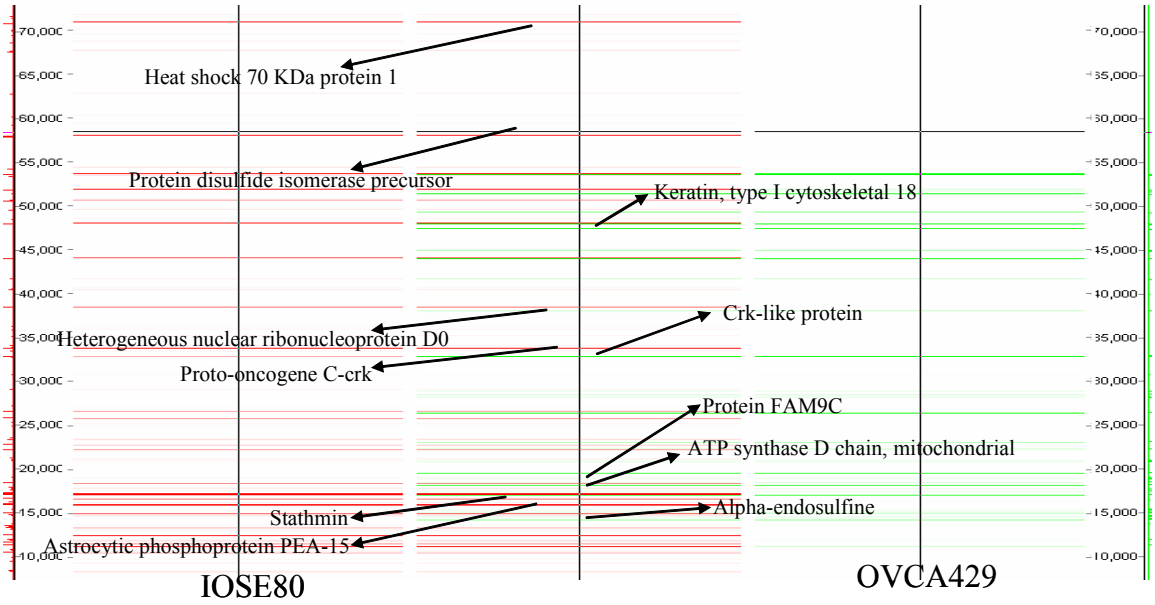


Figure 3.4 A differential display maps between the OVCA429 cell line and IOSE80 from fraction pI 5.15-5.30.

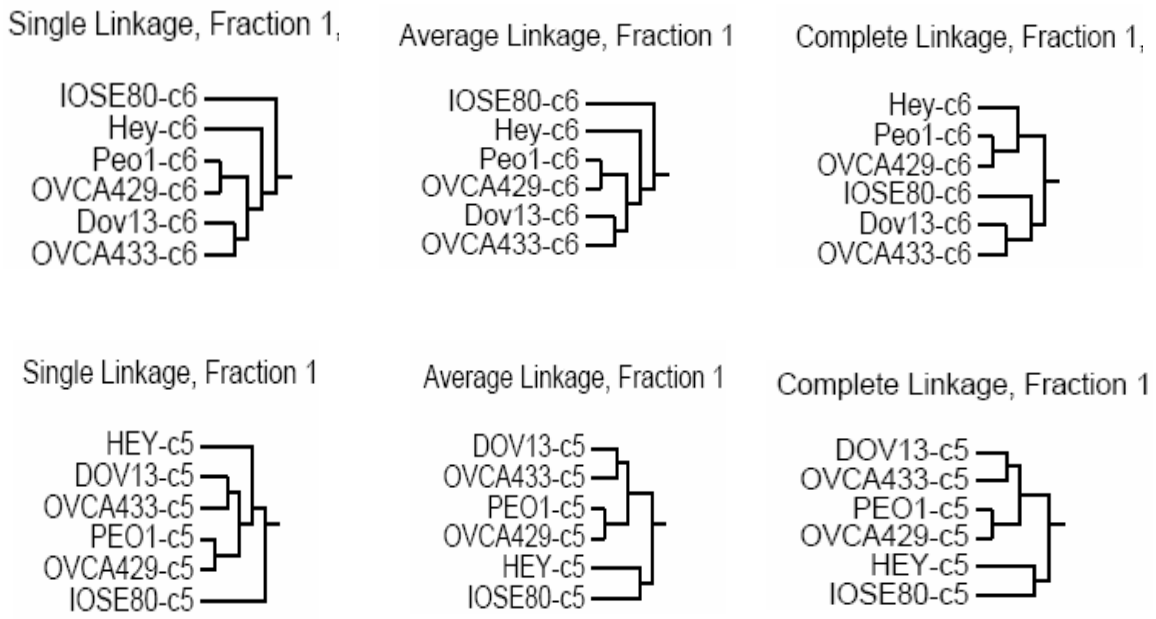


Figure 3.5 Clustering Analysis

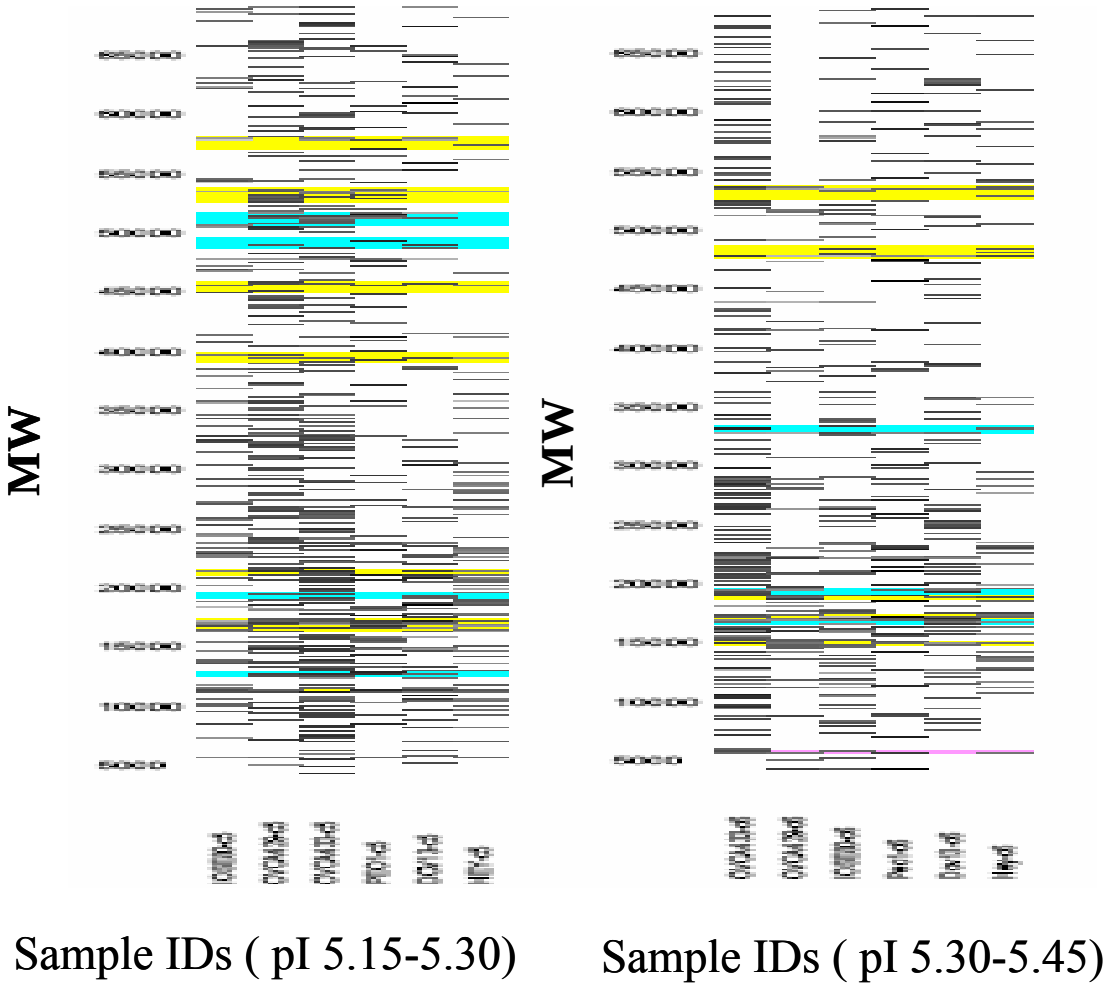


Figure 3.6 An easy view 2-D map for comparison

3.5 References:

1. Ozols, R. F. (2002) Update on the management of ovarian cancer *Cancer Journal*, *8*, S22-S30.
2. Auersperg, N.,Wong, A. S. T.,Choi, K. C.,Kang, S. K., and Leung, P. C. K. (2001) Ovarian surface epithelium: Biology, endocrinology, and pathology *Endocrine Reviews*, *22*, 255-288.
3. Auersperg, N.,Pan, J.,Grove, B. D.,Peterson, T.,Fisher, J.,Maines-Bandiera, S.,Somasiri, A., and Roskelley, C. D. (1999) E-cadherin induces mesenchymal-to-epithelial transition in human ovarian surface epithelium *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 6249-6254.
4. Lin, C. J.,Kohn, E., and Reed, E. (1996) The relationship between borderline ovarian tumors and epithelial ovarian carcinoma: Epidemiologic, pathologic, and molecular aspects *Gynecologic Oncology*, *60*, 347-354.
5. Shih, I. M., and Kurman, R. J. (2004) Ovarian tumorigenesis - A proposed model based on morphological and molecular genetic analysis *American Journal of Pathology*, *164*, 1511-1518.
6. Sawiris, G. P.,Sherman-Baust, C. A.,Becker, K. G.,Cheadle, C.,Teichberg, D., and Morin, P. J. (2002) Development of a highly specialized cDNA array for the study and diagnosis of epithelial ovarian cancer *Cancer Research*, *62*, 2923-2928.
7. Welsh, J. B.,Zarrinkar, P. P.,Sapinoso, L. M.,Kern, S. G.,Behling, C. A.,Monk, B. J.,Lockhart, D. J.,Burger, R. A., and Hampton, G. M. (2001) Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 1176-1181.
8. Ismail, R. S.,Baldwin, R. L.,Fang, J. G.,Browning, D.,Karlan, B. Y.,Gasson, J. C., and Chang, D. D. (2000) Differential gene expression between normal and tumor-derived ovarian epithelial cells *Cancer Research*, *60*, 6744-6749.
9. Hough, C. D.,Sherman-Baust, C. A.,Pizer, E. S.,Montz, F. J.,Im, D. D.,Rosenshein, N. B.,Cho, K. R.,Riggins, G. J., and Morin, P. J. (2000) Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer *Cancer Research*, *60*, 6281-6287.
10. Ye, B.,Cramer, D. W.,Skates, S. J.,Gygi, S. P.,Pratomo, V.,Fu, L. F.,Horick, N. K.,Licklider, L. J.,Schorge, J. O.,Berkowitz, R. S., and Mok, S. C. (2003) Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: Identification and characterization using proteomic profiling and mass spectrometry *Clinical Cancer Research*, *9*, 2904-2911.
11. Wang, H. X.,Kachman, M. T.,Schwartz, D. R.,Cho, K. R., and Lubman, D. M. (2004) Comprehensive proteome analysis of ovarian cancers using liquid phase separation, mass mapping and tandem mass spectrometry: A strategy for identification of candidate cancer biomarkers *Proteomics*, *4*, 2476-2495.
12. Zhou, M.,Prieto, D. A.,Lucas, D. A.,Chan, K. C.,Issaq, H. J.,Veenstra, T. D., and Conrads, T. P. (2006) Identification of the SELDI ProteinChip human serum retentate by microcapillary liquid chromatography-tandem mass spectrometry *Journal of Proteome Research*, *5*, 2207-2216.

13. Binelli, C.,Chretien, M. F.,Yves, G.,Kanaya, M., and Alain, D. (1997) Inhibin assays of ovarian cyst liquid obtained by needle aspiration may allow differential diagnosis between functional and organic cysts *European Journal of Obstetrics Gynecology and Reproductive Biology*, *73*, 161-165.
14. Zhu, Y.,Wu, R.,Sangha, N.,Yoo, C.,Cho, K. R.,Shedden, K. A.,Katabuchi, H., and Lubman, D. M. (2006) Classifications of ovarian cancer tissues by proteomic patterns *Proteomics*, *6*, 5846-5856.
15. Zhu, K.,Kim, J.,Yoo, C.,Miller, F. R., and Lubman, D. M. (2003) High sequence coverage of proteins isolated from liquid separations of breast cancer cells using capillary electrophoresis-time-of-flight MS and MALDI-TOF MS mapping *Analytical Chemistry*, *75*, 6209-6217.

Chapter 4

Differential Protein Mapping of Ovarian Serous Adenocarcinomas: Identification of Potential Markers for Distinct Tumor Stage

4.1 Introduction

Ovarian cancer remains the fifth leading cause of cancer deaths among women in the United States, with an estimated incidence of 22,220 and 16,210 deaths in 2005 (1). Five-year survival rates can be as high as 94% with early detection of the malignancy. However, as ovarian cancer is insidious in onset, less than 20% of ovarian tumors present at early stages of tumor development. The 5-year survival of women with advanced stage ovarian cancer is approximately 28%, largely because existing therapies for widespread disease are rarely curative (2). Currently, the only marker with utility for therapeutic monitoring of ovarian cancer is CA125. However, CA125 does not possess sufficient specificity and sensitivity to have utility for patient screening, given that CA125 is elevated in 52.5% patients with low stage disease (3, 4) and detects only ~80% of all ovarian cancers (5). Thus, additional diagnostic and prognostic markers for ovarian cancer are urgently needed. Identification of novel ovarian cancer biomarkers may lead to the development of efficacious strategies for early detection, as well as improved monitoring of tumor progression and therapy.

Epithelial ovarian cancers (carcinomas) occur as several distinct morphological subtypes, including serous, endometrioid, mucinous, and clear cell carcinomas. Each subtype of ovarian cancer has different clinical characteristics as well as different

molecular features and chemotherapeutic responsiveness (5-9). For ovarian serous carcinomas, a substantial body of literature supports the notion that low grade and high grade OSCs represent distinctly different tumors. In particular, it is believed that the low grade OSCs likely arise from serous borderline tumors, while high grade OSCs likely arise *de novo* (10). Low grade OSCs are typically more indolent than the high grade neoplasms (11, 12), with the low grade neoplasm frequently pursuing a course that may last up to twenty years. Although many patients with low grade OSCs may ultimately succumb due to widespread intra-abdominal spread of the tumor, the tumor maintains its low-grade appearance (13), in contrast to clinically aggressive high-grade neoplasms that spread rapidly and are usually associated with poor outcomes. Molecular studies of OSCs show very distinct genetic alterations between low and high grade neoplasms, with the low grade/well differentiated ones having frequent K-Ras and/or B-Raf mutations in approximately 65% of the tumors and rarely showing p53 mutations, while high grade OSCs show p53 mutations in 50-80% of the cases (14).

Though comprehensive studies of tumor RNA and DNA have provided a number of insights into ovarian cancer pathogenesis, proteins are the major effector molecules in tumor cells. Protein levels may be discordant with corresponding transcript levels and, importantly, a variety of post-translational modifications can have profound biological effects on protein function. To date, a number of proteomics-based studies have been conducted on ovarian tumors or on sera obtained from patients with ovarian tumors using several methods, including 2-D PAGE (15, 16), Surface Enhanced Laser Desorption Ionization (SELDI) (17), and 2-D liquid-based separation methodologies (18, 19). The latter methodology not only has greater reproducibility, but also has the ability to identify

and quantify proteins, and the capability to compare results amongst different sets of experiments and samples (20). Histologic type-specific protein expression patterns in ovarian carcinomas have not been defined to date.

We have compared protein expression in low stage (Stage 1) versus high stage (Stages 3/4) ovarian serous adenocarcinomas (OSCs) to identify stage-dependent protein expression. All but one of the tumors were high grade (grade 2 or 3), thus the observed expression differences between high and low stage tumors was not attributable to differences in the distribution of tumor grade (degree of differentiation) between the two groups. We used a 2-dimensional liquid-based fractionation system to separate and map the protein expression of 19 OSCs from 15 different patients. We utilized hierarchical clustering analysis to explore tumor interrelationships based upon protein expression profiles. The tissues were classified into two groups, with one group consisting of low stage (Stage 1) tumors and another of high stage (Stages 3/4) tumors. Through the use of global protein expression profiling, a number of differentially expressed tumor-associated protein biomarkers with potential biological relevance in ovarian tumorigenesis or tumor progression were identified.

4.2 Materials and Methods

4.2.1 Samples and Preparation

4.2.1.1 Samples

Nineteen snap-frozen primary human ovarian serous carcinoma tissue samples obtained from patients seen at The University of Michigan Hospital were analyzed. Tumors were classified as well differentiated (grade 1), moderately differentiated (grade 2), or poorly differentiated (grade 3) using standard criteria [16, 17]. Tumor stage (1–4)

was assigned according to FIGO criteria. Collection and analysis of human ovarian tumors was approved by the University of Michigan's Institutional Review Board (IRB). The serous ovarian adenocarcinomas were histologically analyzed and staged by a Board-Certified Pathologist (K.R.C.) prior to utilization in this study.

4.2.1.2 Tissue lysis and buffer exchange

All tumors were brought to the laboratory, and then promptly frozen and stored at -80ºC until analysis. Tissue samples were minced into small pieces on dry ice to prevent thawing and subsequently placed into 2-ml screw-cap microcentrifuge vials (BioSpec Products, OK) containing approximately 2/3 vial volume of minute glass beads (BioSpec Products). The vials were subsequently filled with 1ml lysis buffer (7.5 M urea, 2.5 M thiourea, 4% *n*-octyl-β-D-glucopyranoside (*n*-OG), 10 mM tris(2-carboxyethyl) phosphine (TCEP), 12.5% v/v glycerol, and 1% v/v protease inhibitor cocktail (Sigma, St. Louis, MO)). The tissue samples were homogenized for 3 min in a mini-bead beater cell disruptor, followed by centrifugation at 3000 x g for 10 min at room temp to pellet the bead mix. The supernatant containing tumor proteins were collected and stored on ice. In order to avoid incomplete tissue lysis and protein extraction, the vials were filled with 1 ml fresh lysis buffer and again homogenized for 2 min as above. These two lysis solutions were combined in 10 mL polycarbonate centrifuge tubes and the insoluble material was removed by ultracentrifugation at 86,000 x g at 4ºC for 1h. The supernatant was collected and the buffer exchanged with the chromatofocusing start buffer using a PD-10 G-25 column (Amersham Biosciences, Piscataway, NJ). The proteins were quantified with the Bradford assay (Bio-Rad), and then the buffer-exchanged protein mixtures were stored in a -80ºC freezer until further use.

4.2.2 Separation

4.2.2.1 Chromatofocusing

CF separation was performed on an HPCF-1D column (250 × 2.1 mm) (Beckman Coulter, Fullerton, CA) using the ProteomeLab™ PF2D protein fractionation system (Beckman Coulter), as described previously [18]. Two buffers were used to generate the pH gradient on the column. The SB solution was composed of 6M urea, 25mM Bis-Tris (pH 7.4). The EB solution was composed of 6M urea and 10% polybuffer74 (pH 4.0). Both buffer solutions were brought to pH by addition of a saturated solution of iminodiacetic acid. The CF column was pre-equilibrated with SB. After equilibration, 4.5 mg of proteins were loaded onto the CF column and the column was washed with 100% SB to remove material that did not bind to the column at pH 7.4. Elution is achieved by applying a pH 4.0 elution buffer at a flow rate of 0.2 mL/min. The pH gradient was monitored on-line by a flow-through pH probe (Beckman Coulter). The UV absorbance of the eluent was monitored on-line at 280nm. The flow rate was 0.2ml/min, with 16 fractions in total being collected in 0.2 pH units in the range of pH 7.0 - 4.0. Each fraction was stored at -80°C until use.

4.2.2.2 NPS-RP-HPLC with sample collection

When the first-dimension separation was completed, the pI fractions collected from the first dimension were then separated by nonporous silica reverse phase HPLC (NPS-RP-HPLC) using a HPCF-2D (4.6 × 33 mm) NPS column (Beckman Coulter) and detected by absorbance at 214 nm using a Beckman model 166 UV absorption detector. The RP separation was performed at 0.75 mL/min and monitored at 214 nm using a Beckman 166 Model UV detector (Beckman-Coulter). Proteins eluting from the column

were collected by an automated fraction collector (Model SC 100, Beckman), controlled by an in-house designed DOS-based software program. To enhance the speed, resolution, and reproducibility of the separation, the RP column was heated to 65ºC by a column heater (Jones Chromatography, Model 7971, Resolution Systems, Holland, MI). Both mobile phase A: MilliQ® water (Millipore, Billerica, MA), and solvent B: acetonitrile (ACN) (Sigma) contains 0.1% v/v trifluoroacetic acid (TFA). The gradient was run from 15 to 25% B in 1 min, 25 to 35% in 6 min, 35 to 38% in 4 min, 38 to 45% in 6 min, 45 to 65% in 2 min, 65 to 67% in 6 min, and finally up to 100% in 1 min, then back to 5% in 1 min. After the gradient, the column was washed by two fast gradients from 5% B to 100% B in 5 min, 100% B back to 5% B in 1 min. Fractions from the HPLC eluent were collected using a semi-automated in-house program using a Model SC-100 fraction collector. Collected peak fractions were stored at -80ºC for further use.

4.2.3 Protein Identification

4.2.3.1 Tryptic Digestion Nonporous Reversed-Phase HPLC Fractions

Second dimension fractions of interest (obtained from NPS-RP-HPLC) were concentrated to 20 μL with a SpeedVac concentrator (Labconco, Kansas City, MO) operating at 45ºC. Twenty microliters of 200 mM ammonium bicarbonate (Sigma) and 2 μL of 100 mM dithiothreitol (DTT) (Sigma) were mixed with each concentrated sample to obtain a pH value of approximately 8.0. 0.5μL of L-1-tosylamido-2-phenylethyl chloromethylketone (TPCK) modified sequencing-grade porcine trypsin (Promega, Madison, WI) was vortexed prior to an overnight incubation at 37ºC, with agitation. Following incubation, 1μl TFA was added to each tube to halt digestion.

4.2.3.2 MALDI-TOF-MS Sample Preparation and Data Acquisition

Before MALDI analysis, each tryptic digested sample was desalted and concentrated using C18 ZipTip (Millipore) into 5 μL 60% (v/v) ACN/0.1% (v/v) TFA. One microliter of the concentrated peptide mixture was spotted onto a Micromass 96-spot plate followed by 1 μL of matrix-standard layered on top of it. The MALDI matrix solution was prepared by diluting saturated α-CHCA (Sigma) solution with 60% v/v ACN and 0.1% v/v TFA at 1:4 ratio v/v with three internal standard proteins: angiotensin I (1296 Da), adrenocorticotropic hormone (ACTH clip 1–17, 2093 Da), and ACTH clip 18–39 (2465 Da) (Sigma). 1 mg/mL standard proteins were diluted 100-fold with deionized water. These standards were further diluted 50, 40, and 30 times, respectively, with the diluted matrix.

The Micromass TofSpec-2E™ (Waters) was used to generate peptide mass fingerprints (PMF), followed by database searching in the SwissProt database 10.30.2003. Spectra were calibrated using the internal calibrants to achieve a mass accuracy of less than 50 ppm. Monoisotopic peak lists were generated using Masslynx 4.0 and searched against the SwissProt protein database, using the MS-Fit search engine 10/20/2006 (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm). For the peak picking, we typically adjust the data threshold so the peaks will be around 60 to 100 for each spectrum with the internal standard peak excluded. The search was performed using the following parameters: (1) species: human; (2) allowing one missed cleavage; (3) possible modifications: peptide *N*-terminal glutamine to pyroglutamic acid, oxidation of methionine, and protein *N*-terminus acetylated and phosphorylation of S, T and Y; (4) peptide mass tolerance 50 ppm; (5) MW ranged from 1000 to 100,000 Da; (6) p*I* range of protein 3–10. Protein identifications were considered confident matches when meeting

the following criteria: (i) ranked in the top three database hits; (ii) sequence coverage is >20%; (iii) average mass error is <50 ppm (iv) MOWSE scores over 10^4.

4.2.3.3 QIT MALDI quadrupole ion trap-ToF Sample preparation and Data Acquisition

Before MALDI analysis, each tryptic digested sample was desalted and concentrated using C18 ZipTip (Millipore) into 5 μL 60% (v/v) ACN/0.1% (v/v) TFA. 0.5 microliter of the concentrated peptide mixture was spotted onto a steel plate followed by 0.5 μL of matrix-standard layered on top of it. The MALDI matrix solution was prepared as 20 mg/mL 2,5-dihydroxybenzonic acid (DHB) (LaserBio Labs, France) in 50% ACN with 0.1% TFA. A pulsed N_2 laser light (337 nm) with a pulse rate of 5 Hz was used for ionization. Each profile combines 2 laser shots. Argon was used as the collision gas for CID and helium was used for cooling the trapped ions. The TOF was externally calibrated using 1 pmol/μL of bradykinin fragment 1-7(757.40 *m*/*z*), angiotensin II (1046.54 *m*/*z*), P14R (1533.86 *m*/*z*) and 2 pmol/μL ACTH (2465.20 *m*/*z*) (sigma). Data were acquired on a Shimadzu Axima (MALDIQIT) (Manchester, UK). Acquisition and data processing were controlled by Launch-pad software V2.4.1 (Karatos, Manchester, UK). Monoisotopic peak lists were submitted to Swiss-Prot protein database 2006.06.20, using Mascot/peptide mass fingerprint search engine version 2.1 (http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF).

The search was performed using the following parameters: (1) species: human; (2) trypsin; (3) allowing one missed cleavage; (4) peptide mass tolerance 1.2 Da, MS/MS tolerance 0.8Da; Protein identifications are considered confident matches when meeting

the following criteria: (1) Individual ions scores indicate the peptide identity or extensive homology; (2) At least two peptides matching.

4.2.3.4 LC-MS/MS

The tryptic-digested samples were separated by a capillary RP column (C18, 0.3 × 150 mm) (Michrom Biosciences, Auburn, CA) on a Paradigm MG4 micropump (Michrom Biosciences) with a flow rate of 5 μL/min. The gradient, started at 5% ACN, was ramped to 60% ACN in 25 min and finally ramped to 95% in another 5 min. Both solvents A (water) and B (ACN) contain 0.3% formic acid. The resolved peptides were analyzed on a Finnigan LTQ mass spectrometer (Thermo Electron Corp., San Jose, CA) with an ESI ion source (Thermo). The capillary temperature was set at 175°C, spray voltage was 2.8 kV, and capillary voltage was 30 V. The normalized collision energy was set at 35% for MS/MS. MS/MS spectra were searched using the SEQUEST algorithm, Version 27, incorporated in Bioworks software Rev. 3.1 SR1(Thermo) against the Swiss-Prot human protein database 2007.01.09. The search was performed using the following parameters: (1) database species, *Homo sapiens*; (2) allowing one missed cleavages; (3) possible modifications, none; (4) peptide ion mass tolerance 1.50 Da; (5) fragment ion mass tolerance 0.0 Da; (6) mass tolerance for precursor ions 1.40 Da; (7) peptide charges +1, +2, and +3. Protein identification was considered positive for a peptide with *X*corr of greater than or equal to 3.5 for triply, 2.5 for doubly, and 1.9 for singly charged ions. While no ions at higher charged states were considered. The search results that passed the criteria were subjected to be considered when $\Delta Cn \geq 0.1$. If multiple members of a protein family were identified, criteria used for selecting were as follows: (1) ranked preliminary score (Rsp) <5; (2) ions >70%; (3) at least 2 peptides matching; (4) if the

same spectrum matched different proteins, the higher Rsp, the higher ΔCn, and/or ions >70% would be selected.

4.2.4 Software

The data from the 2-D liquid separations were displayed using the Mapping Tools Software Suite from Beckman-Coulter. The chromatographic UV data resulting from the NPS-RP-HPLC second-dimension separation of each pI fraction were analyzed by the ProteomeLab software to produce peak information (peak retention time, peak height, peak area). The Mapping Tools Software Suite imports the raw chromatogram intensity vs. retention time data and those processed peaks for each pI range in a data set. The multiple data sets (one for each pI range as a "lane") were displayed as a two dimensional image based upon retention time versus pI range axis (termed a ProteoVue map). Each raw chromatogram intensity and/or processed peak data value is shown proportional to color intensity or color value at its corresponding retention time and pI range location.

Two such complete data sets or maps were compared. Corresponding peaks in similar pI ranges were selected for comparison. A DeltaVue difference map was prepared from the paired peaks. A single pI range can be compared across many data sets by creating a lane-band pattern image map. One data set was defined as the Reference data set and the peaks in the Reference lane were paired with peaks in each of the individual Comparison data set's lane. A MultiVue map was prepared for these paired peaks to allow comparison of many data sets at a specific pI value.

4.2.5 Data analysis and clustering

Data Standardization and Alignment:

The raw UV data for each sample were standardized to remove differences in the level and slope of the baseline. After standardization, the UV data for each pair of samples were aligned in order to maximize the local correlation coefficients between the aligned samples. The method used for the standardization and alignment was identical to that utilized previously [18].

Comparisons:

To compare the overall pattern of protein expression in the samples, each pair was aligned separately, and then a correlation matrix was formed by calculating the Pearson correlation coefficient between each aligned pair of samples. These correlation matrices were then visualized using a hierarchical clustering technique, thus producing a dendrogram in which pairs of points with greater correlation are linked sooner. Complete linkage clustering analysis was used to define the dendrograms.

4.3 Results and Discussion

4.3.1 Analysis of ovarian serous carcinoma samples

Nineteen OSC tumor samples were used for proteomic analysis. Available clinico-pathological data associated with tumor specimens is provided in Table 1. Notably, all but one of the tumors were high grade (grade 2 or 3), so that expression differences between high and low stage tumors was not attributable to differences in the distribution of tumor grade (degree of differentiation) between the two groups. Figure 1 provides an overview of the experimental procedure. 2-D liquid-based mapping methodology was developed and applied to profile global protein expression of both low stage (stage 1) and high stage (stage 3/4) OSCs. A hierarchical clustering technique was

applied to classify the samples which were followed by the identification of differentially expressed proteins of different groups based on the classification.

4.3.2 Proteome analysis

In this study, the global protein expression profiles for 19 OSC tissue samples (7 stage 1 (from 5 individuals) and 12 stage 3/4 (from 10 individuals)) were generated using 2-D liquid mapping methodology to assess protein expression differences between both low and high stage OSCs. Maps were produced using protein pI as the separation parameter in the first dimension and hydrophobicity, based upon reverse phase (RP)-HPLC separation, in the second dimension. Protein detection in the first dimension was performed using UV absorption at 280 nm, with UV absorption at 214 nm in the second dimension. Equal amounts of protein were loaded for each sample in order to generate the 2-D UV maps. As shown (Figure 2), 4.5 mg of tumor proteins were loaded onto the first dimension chromatofocusing column. In this figure each lane corresponds to a different pI value, and the bands correspond to the hydrophobicity as generated by the percentage of acetonitrile on the HPLC gradient at that pI. Each map consists of a total of 15 pI fractions, corresponding to a pH range of 4.0-7.0.

The chromatographic UV data resulting from the nonporous silica RP-HPLC second dimension separation of each pI fraction were displayed using ProteoVue software and analyzed to obtain peak information. The processed peaks were displayed in a 2-D “lane and band” format resulting in a highly detailed pI versus hydrophobicity protein expression map, using a color-coded format where color hue or its intensity is proportional to the relative quantitative UV peak volume of each peak, thus facilitating viewing of relationships or patterns within the complex chromatographic data set.

4.3.3 Reproducibility studies

The reproducibility of the analysis technique was studied. Figure 3A represents the first dimension chromatofocusing profile for sample UM-OS-21-1, UM-OS-21-2 and UM-OS-21-3. Sample UM-OS-21-1 and UM-OS-21-2 are different protein lysates from the same tumor that were utilized for analysis of tumor heterogeneity. Sample UM-OS-21-3 is the same sample as UM-OS-21-1, and was run twice for the first dimension chromatofocusing to explore experimental reproducibility. The second dimension hydrophobicity profiles for all pI fractions are shown in Fig. 3B. These results indicate high reproducibility between the chromatograms of the band patterns and retention times in both the first dimension chromatofocusing separation and the second dimension reverse phase liquid separation. Such reproducibility was observed in all the pI lanes separated for all three samples (data not shown).

4.3.4 Hierarchical Clustering

Complete linkage cluster analysis was performed following standardization and alignment. All pI fractions, from pH 4.0-7.0, were analyzed by this method. Total results were obtained by the average of all the fractions, and are shown in Figure 4A.

In the dendrograms in Figure 4, the length and the subdivision of the branches display the relatedness of the cell lines and the expression of the proteins. The dendrogram illustrated in Figure 4A shows a major division in the distribution of different OSC tissue samples based upon respective protein expression using all protein fractions generated during the 2-D liquid mapping. In the rightmost major node (black dots), seven stage 1 OSC samples clustered together, with the three independent replicate OSC samples (UM-OS-021(1-3)) from the same patient clustering on the same node. In

the leftmost major node, the twelve stage 3/4 OSC tumor samples clustered together. Two independent sets of replicate tumor samples (UM-OS-7T/UM-OS-7Tb and UM-OS-014T (met 1 and met 2)) were found to cluster tightly together with the corresponding replicate OSC sample. However, upon analysis of individual fractions, differences in clustering patterns were noted. Although all of the replicate OSC samples clustered tightly with the corresponding replicate sample from the same patient, thus indicating good reproducibility of the methodology, the difference between different stages was very apparent in the low pH fractions. Figure 4B represents the pH fraction from 4.4-4.6, with all the low stage tumors being grouped together, apart from the high stage tumors. The low and high stage tumors did not completely segregate in the pH fraction from 6.0-6.2 based on their protein expression profiles (Figure 4C). Within this pH range, the expression profiles of five low stage OSCs were clustered with one high stage tumor (UM-OS-017Tb). All other high stage tumors clustered with two low stage tumors (CHTN-OS-041 and CHTN-OS-050). Thus, whereas protein profiling of all fractions simultaneously provided good segregation of high and low stage OSCs, protein profiling of individual fractions indicated that some low and high stage OSCs may have a similar protein expression profile.

4.3.5 Stage-dependent Candidate Protein Markers of Ovarian Serous Carcinoma

We next sought to identify proteins whose expression pattern could serve as potential stage-dependent biomarkers of OSC. In these comparisons between different stage samples, only peak areas that demonstrated at least 2-fold changes in quantitative protein expression were selected, and analyzed further by mass spectrometry. To individually identify the proteins in each peak, most of the proteins were identified using

LTQ analysis, with the data shown in Figure 5. Additionally, some protein ID's were confirmed using a MALDI-TOF MS or combined with QIT-MALDI-MS analysis (Data not shown. Peptide mapping fingerprints (PMF) were used for protein identification. The differentially expressed proteins that were identified are shown in Figure 5. Each cell in the matrix represents the expression level of a single protein in a single sample, with red or orange indicating intensity above the medium and blue or dark blue indicating intensity below the median for that protein across all samples.

$$\text{Ratio} = \frac{\text{expression level of a single protein in stage 1}}{\text{average expression level of proteins in stage 3 / 4}} \quad \text{or}$$

$$\text{Ratio} = \frac{\text{expression level of a single protein in stage 3 / 4}}{\text{average expression level of proteins in stage 1}}$$

Upon protein identification, we found that more than one protein was confidently identified for some peaks. In these cases, a spectral counting-based quantitation method (24) was used. Using this method, we split the protein expression levels that were obtained from mapping tools between the identified proteins, based on the ratio of their quantitation. We also performed off-line ESI mass spectrometry on the LCT for sample UM-OS-007 fraction 4.6-4.4, with a peak retention time of 11 min to confirm the spectral counting-based quantitation method. As shown in Figure 6a, the proteins contained in this fraction were identified as Tumor Protein D54 and astrocytic phosphoprotein PEA-15. After spectrum deconvolution, we observed that the intensity of Tumor Protein D54 was 1.71×10^2 and for astrocytic phosphoprotein PEA-15 the intensity was 1.42×10^2. For the spectral counting-based quantitation method, the numbers were 1189 and 1166,

respectively. Thus, the total expression levels were split for these two proteins based on the protein-protein ratio of 1.20.

A subset of differentially expressed proteins was further evaluated as possible candidate biomarkers to distinguish between different stages in OSC samples. We focused on several proteins previously associated with cancer that displayed the largest-fold change. The image in Figure 6A is displayed in a format with each different sample on the x-axis and hydrophobicity on the y-axis. The relative intensities of the peak are quantitatively proportional to the amount of corresponding protein detected by UV absorption. The leftmost six columns are samples from low stage tumors, while the rightmost twelve columns are samples from high stage tumors. The band circled in the picture shows high expression in the high stage tumors, but almost no expression in low stage tumors, and has been identified as containing both astrocytic phosphoprotein PEA-15 and Tumor Protein D54. Tumor Protein D54 belongs to the *TPD52* family, which consists of four members, *TPD52* or *D52*, *TPD52L1* or *D53*, and *TPD52L2* or *D54*. The *Tumor Protein D52* (*TPD52*) gene was originally identified as being over-expressed in human breast carcinomas specifically within the neoplastic cells and was subsequently shown to encode a tumor-associated antigen in breast carcinoma through SEREX screening (25). Tumor Protein D52 is known to be over-expressed in breast, prostate and ovarian carcinomas (26,(27,(28,(29). Byrne et al (23) demonstrated that normal ovarian epithelium samples were predominantly TPD52-negative, whereas TPD52 was over-expressed in most (44/57; 77%) ovarian carcinomas regardless of histological subtype. The encoded proteins are likely to function as adaptor proteins and to interact with D52-like proteins and several other intracellular partner proteins (30). We have shown that

Tumor Protein D54 is over-expressed in high stage OSC tumors, as compared to low stage tumors (Figure 6B). Figure 6C shows the tandem mass spectra of Tumor Protein D54-specific peptides.

Vimentin, an intermediate filament protein, was significantly over-expressed in the high stage tumors as compared with the low stage tumors in our study (Figure 7A, 7B). It is expressed in many hormone-independent mammary carcinoma cell lines. Vimentin is found in various cell types, especially mesenchymal cells. Previous reports have also demonstrated the increase of vimentin in the hyperplastic tissue compared to adjacent normal human prostatic epithelium (31). Figure 7A shows vimentin expression (as indicated by UV absorption) and Figure 7B indicates protein expression levels for each of the individual tumors, sorted by stage. Figure 7C shows the tandem mass spectra of vimentin-specific peptides.

Histone proteins were also found to be differentially expressed amongst low and high stage OSCs. Histone proteins are core components of nucleosome which wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. Nucleosome remodeling is regulated via a complex set of post-translational modifications of histones. Acetylation of the lysine residues at the N-terminus of histone proteins removes positive charges, thus reducing the affinity between histones and DNA, and allowing for RNA polymerase and transcription factors to access the promoter region (Figure 8B). Thus, histone acetylation enhances transcription while histone deacetylation represses transcription. We have found many histone proteins over-expressed in the high stage

tumors relative to low-stage tumors. Typically, histones have a very basic pI, usually above 11. As we have found them with pI's of less than 7, this suggests that these proteins may be heavily acetylated in the high stage tumors, thus shifting the pI of those proteins into our detection range, and which may suggest enhanced transcription in the high stage tumors. Figure 8A shows the tandem mass spectra of a histone peptide with a two acetylation modification.

4.3.6 Function of differentially expressed proteins

A biological process clustering of differentially expressed proteins in low- versus high stage OSCs was created to analyze the distribution of important cellular regulatory functions. Based on a deregulation ratio threshold of +/- 2, we found 49 proteins over-expressed in the high stage OSCs (Figure 9A) and 14 under-expressed proteins (Figure 9B), with respect to the low stage tumors. Rather than being evenly distributed in all protein function categories, the over- and under-expressed proteins were clustered within specific biological processes. The percentage of identified proteins in each assigned molecular function was calculated based on the number of assigned proteins over the total proteins. For the proteins that were over-expressed in the high stage tumors, 28% of identified proteins are involved in cytoskeletal organization, 19% are involved in transcription/translation, and 13%, 13%, 6%, 4% and 2% (respectively) are involved in cellular metabolism, response to stress, antioxidant, signal transduction or are mitochondria-related. 15% of the identified proteins have been classified as other (or unknown) function. For the under-expressed proteins, 0%, 7%, 33%, 7%, 0%, 20% and 20% (respectively) are involved in cytoskeletal organization, transcription/translation, cellular metabolism, response to stress, antioxidant, signal transduction or mitochondria-

related. Proteins with other (or unknown) function are 13%. Notably, proteins involved in cytoskeleton organization or antioxidant pathways were found to be over-expressed in the high stage tumors relative to their proportion in the differentially expressed protein population. Interestingly, the mitochondria-related proteins that we have identified are clearly under-expressed in the high stage tumors, consistent with the study of Cuezva et al (32) who showed that liver carcinogenesis involves a depletion of the cellular mitochondrial content, as revealed by reduced content of mitochondrial markers.

Finally, of interest, the low stage OSC samples showed more protein in the acidic fractions and less protein in the basic fractions, as compared to high stage OSC samples. This phenomenon may result not only from changes in gene/protein expression, but also from alterations in the levels and types of post-translational modifications of expressed proteins (i.e., changes in acetylation and glycosylation, etc.), changes that are more difficult to ascertain on an individual basis, but can be observed globally by protein profiling. Thus, progress to define markers for diagnosis of low stage OSC may be facilitated by the establishment of protein profiles rather than by analysis of uniquely-expressed proteins.

4.4 Conclusion:

In conclusion, we utilized 2-D liquid mapping and hierarchical clustering analysis of 19 OSC tissue samples (seven stage 1 (from 5 individuals) and twelve stage 3/4 (from 10 individuals). Replicate samples obtained from the same tumor demonstrated good reproducibility in our analyses. Further, we have demonstrated that the tumors could be classified in two groups, with one group associated with low stage OSCs and the others associated with high stage tumors. Notably, all but one of the tumors were high grade

(grade 2 or 3), so that expression differences between high and low stage tumors was not attributable to differences in the distribution of tumor grade (degree of differentiation) between the two groups. To our knowledge such a systematic large-scale proteomic study has not before been achieved. With differential expression maps, 64 proteins were classified as over-expressed or under-expressed between the low and high stage OSCs. Further analysis of the differentially expressed proteins may facilitate identification of candidate prognostic markers in ovarian cancer.

Tumor ID	Stage	Grade	Age
UM-OS-001T	3	2	70
UM-OS-002T	3c	2-3	40
UM-OS-003T	4	3	56
UM-OS-007T	3c	3	54
UM-OS-007Tb	3c	3	54
UM-OS-014 (met 1)	3c	3	54
UM-OS-014 (met 2)	3c	3	54
UM-OS-015T	4	3	49
UM-OS-016T	3c	1	30
UM-OS-017Tb	3c	2-3	81
UM-OS-019T	3	3	44
UM-OS-020T	3	3	65
UM-OS-021-1T	1a	3	68
UM-OS-021-2T	1a	3	68
UM-OS-021-3T	1a	3	68
CHTN-OS-004	1a	2	47
CHTN-OS-041	1a	3	61
CHTN-OS-050	1c	3	80
CHTN-OS-098	1a	3	59

Table 4.1 Serous ovarian carcinoma tumors utilized for proteomic analysis of ovarian cancer proteins

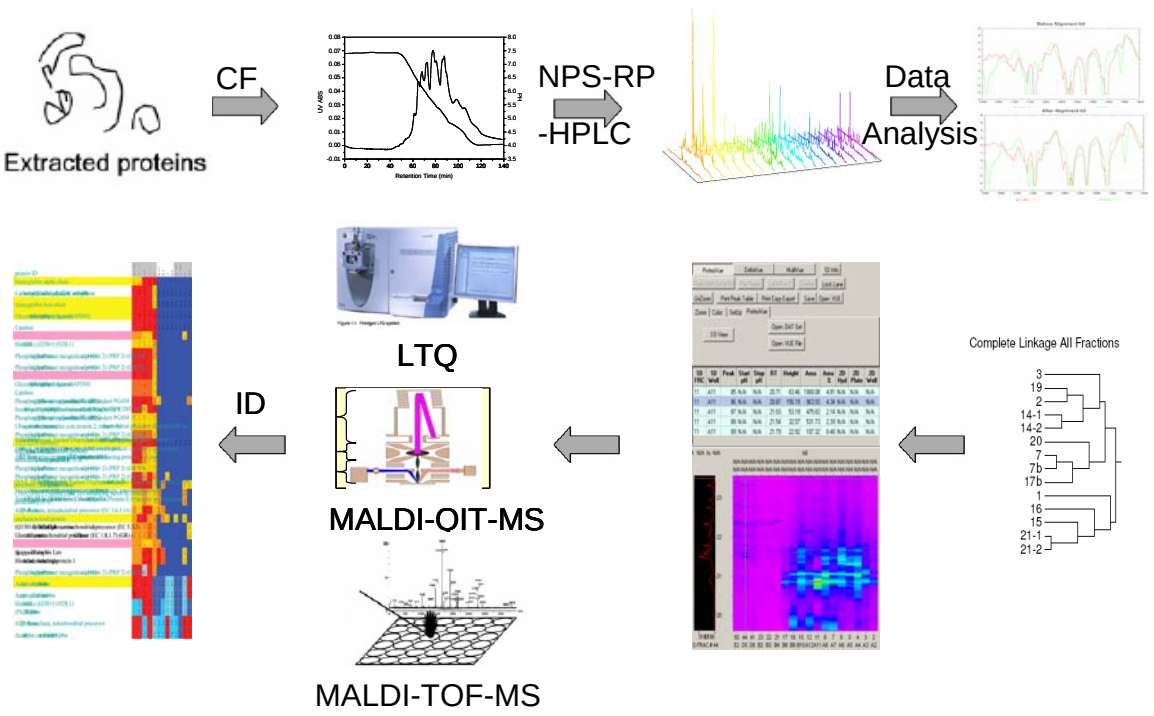


Figure 4.1 Experimental flow chart for global protein expression profiling maps. Tissue samples were lysed and the soluble proteins separated by pI in the first dimension using chromatofocusing, and by NPS-RP-HPLC in the second dimension. Data analysis was applied to the raw data and samples were grouped based on similarities in their protein expression using a hierarchical clustering analysis technique. Differential proteins were identified by several mass spectrometry methods.

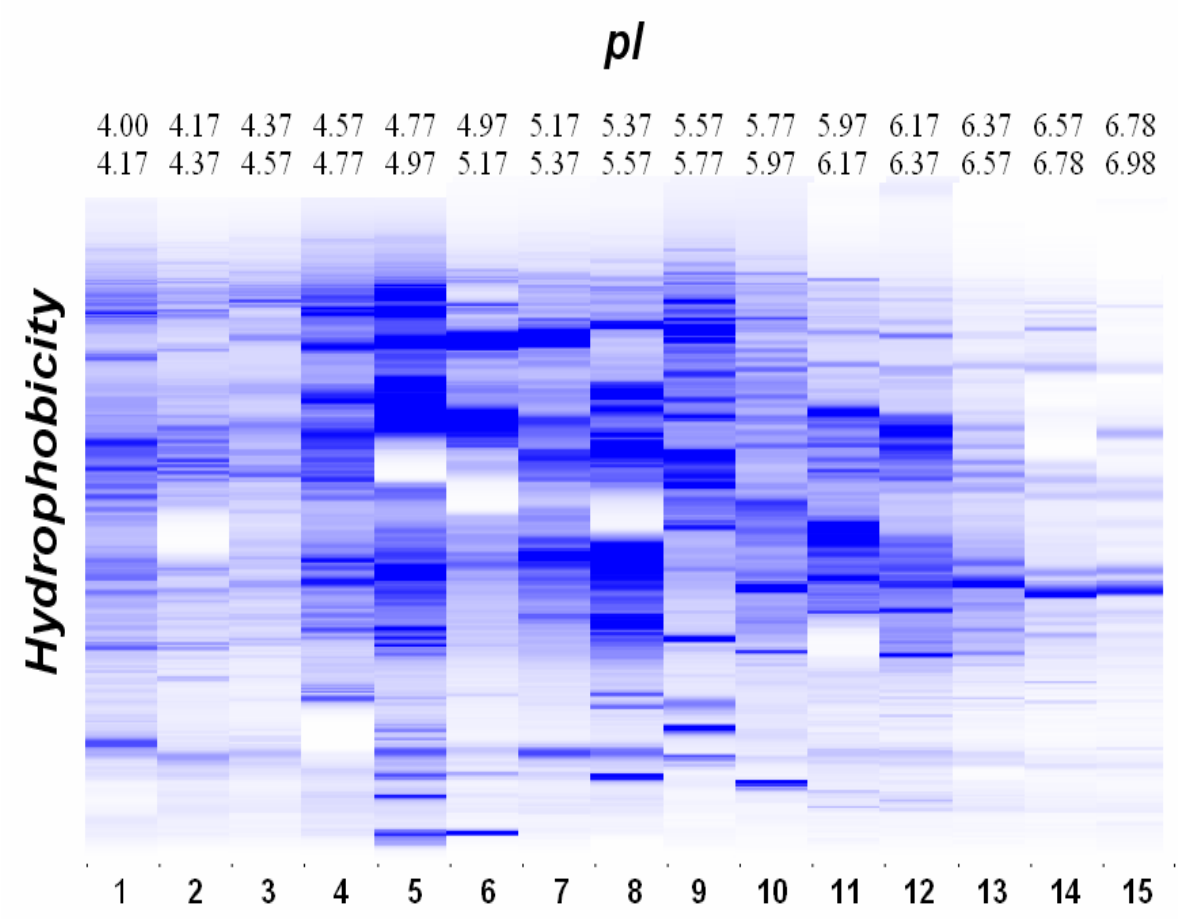


Figure 4.2 Ovarian cancer serous carcinoma tissue sample UM-OS-007 was separated using chromatofocusing over a pH range of 4.0–7.0, in 0.2 pH intervals, followed by separation in the second dimension using NPS RP-HPLC. The x-axis is pI of the chromatofocusing and the y-axis is hydrophobicity of the RP-HPLC. The scale of the bands represents the relative intensity of each band by UV detection at 214nm.

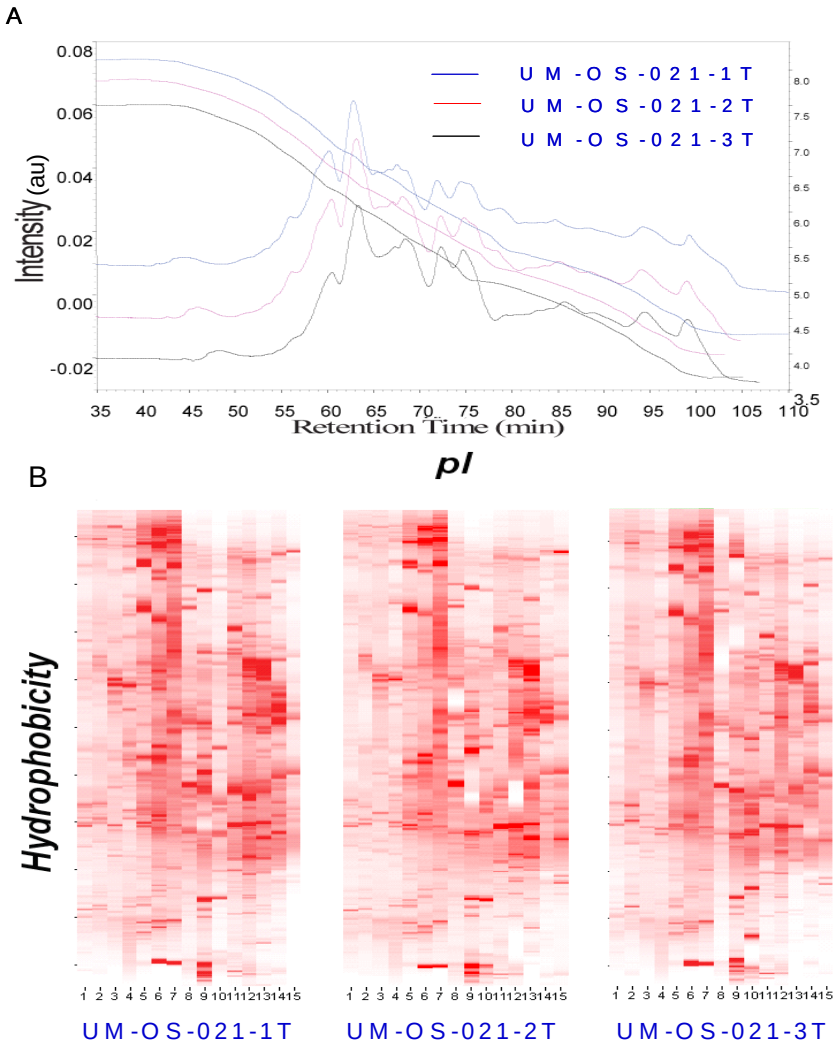


Figure 4.3 A. Analysis of technical reproducibility. Displayed is the first dimension chromatofocusing profile for sample UM-OS-21-1, UM-OS-21-2 and UM-OS-21-3. Sample UM-OS-21-1 and UM-OS-21-2 are different tissue samples from the same tumor that were used to study tumor heterogeneity. Sample UM-OS-21-3 is the same sample as UM-OS-21-1, and was run twice for chromatofocusing to explore experimental reproducibility. Total amount of each sample was 4.5mg, and the protein content of the samples was detected by UV absorption at 280nm. B. 2nd dimension separation was run continuously by NPS RP-HPLC for each CF separation. UV maps are shown. UV absorption is at 214nm.

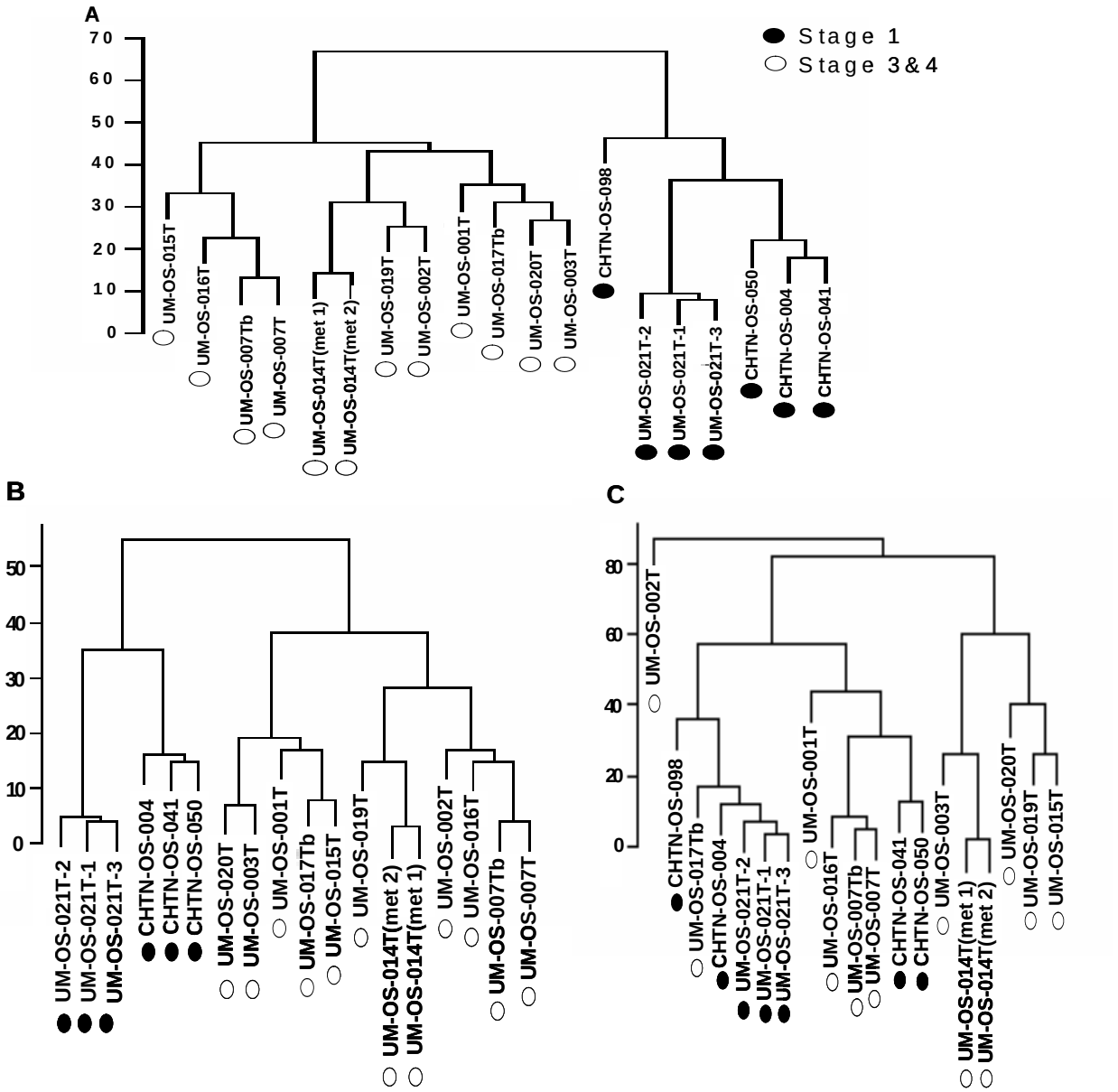
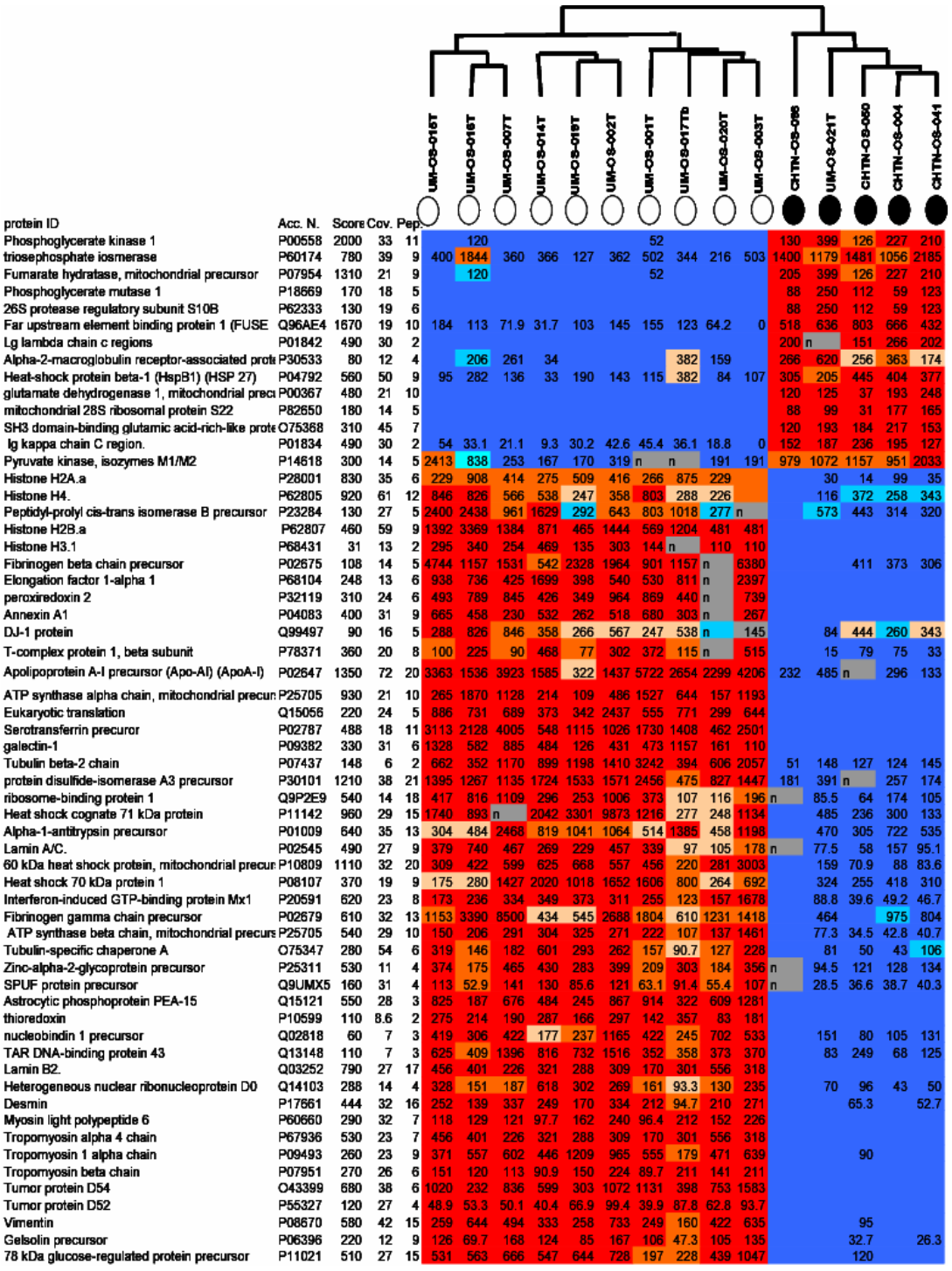


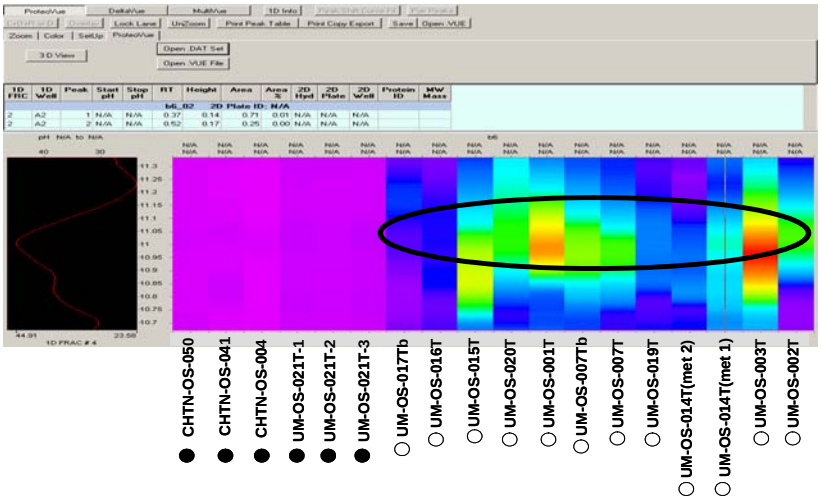
Figure 4.4 Tissue samples were aligned separately and grouped based on similarities in their protein expression using complete linkage hierarchical clustering analysis. This technique produces a dendrogram in which pairs of points are joined sooner (i.e. closer to the ends of the dendrogram) if they have greater correlation. A. Dendrogram of samples showing overall similarity in protein expression profiles across the samples. B. pI fraction 4.4-4.6 protein expression profiles across the samples C. pI fraction 6.0-6.2 protein expression profiles across the samples.



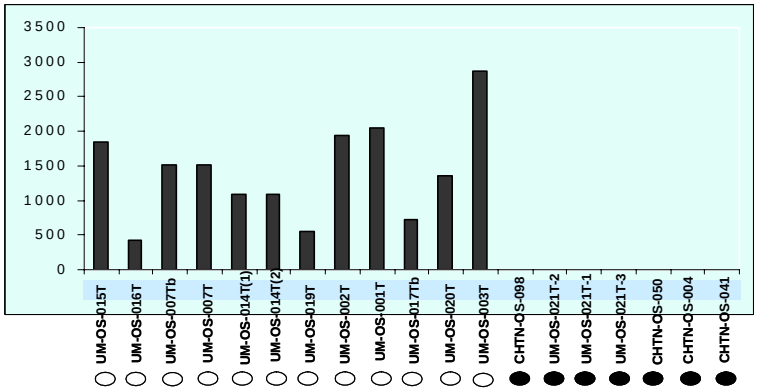
protein ID	Acc. N.	Score	Cov.	Pep.															
Phosphoglycerate kinase 1	P00558	2000	33	11		120					52				130	399	126	227	210
triosephosphate isomerase	P60174	780	39	9	400	1844	360	366	127	362	502	344	216	503	1400	1179	1481	1056	2185
Fumarate hydratase, mitochondrial precursor	P07954	1310	21	9		120					52				205	399	126	227	210
Phosphoglycerate mutase 1	P18669	170	18	5											88	250	112	59	123
26S protease regulatory subunit S10B	P62333	130	19	6											88	250	112	59	123
Far upstream element binding protein 1 (FUSE	Q96AE4	1670	19	10	184	113	71.9	31.7	103	145	155	123	64.2	0	518	636	803	666	432
Lg lambda chain c regions	P01842	490	30	2											200	n	151	266	202
Alpha-2-macroglobulin receptor-associated prot	P30533	80	12	4		206	261	34				382	159		266	620	256	363	174
Heat-shock protein beta-1 (HspB1) (HSP 27)	P04792	560	50	9	95	282	136	33	190	143	115	382	84	107	305	205	445	404	377
glutamate dehydrogenase 1, mitochondrial preci	P00367	480	21	10											120	125	37	193	248
mitochondrial 28S ribosomal protein S22	P82650	180	14	5											88	99	31	177	165
SH3 domain-binding glutamic acid-rich-like prote	O75368	310	45	7											120	193	184	217	153
Ig kappa chain C region.	P01834	490	30	2	54	33.1	21.1	9.3	30.2	42.6	45.4	36.1	18.8	0	152	187	236	195	127
Pyruvate kinase, isozymes M1/M2	P14618	300	14	5	2413	838	253	167	170	319	n	n	191	191	979	1072	1157	951	2033
Histone H2A.a	P28001	830	35	6	229	908	414	275	509	416	266	875	229			30	14	99	35
Histone H4.	P62805	920	61	12	846	826	566	538	247	358	803	288	226			116	372	258	343
Peptidyl-prolyl cis-trans isomerase B precursor	P23284	130	27	5	2400	2438	961	1629	292	643	803	1018	277	n		573	443	314	320
Histone H2B.a	P62807	460	59	9	1392	3369	1384	871	465	1444	569	1204	481	481					
Histone H3.1	P68431	31	13	2	295	340	254	469	135	303	144	n	110	110					
Fibrinogen beta chain precursor	P02675	108	14	5	4744	1157	1531	542	2328	1964	901	1157	n	6380			411	373	306
Elongation factor 1-alpha 1	P68104	248	13	6	938	736	425	1699	398	540	530	811	n	2397					
peroxiredoxin 2	P32119	310	24	6	493	789	845	426	349	964	869	440	n	739					
Annexin A1	P04083	400	31	9	665	458	230	532	262	518	680	303	n	267					
DJ-1 protein	Q99497	90	16	5	288	826	846	358	266	567	247	538	n	145		84	444	260	343
T-complex protein 1, beta subunit	P78371	360	20	8	100	225	90	468	77	302	372	115	n	515		15	79	75	33
Apolipoprotein A-I precursor (Apo-AI) (ApoA-I)	P02647	1350	72	20	3363	1536	3923	1585	322	1437	5722	2654	2299	4206	232	485	n	296	133
ATP synthase alpha chain, mitochondrial precur	P25705	930	21	10	265	1870	1128	214	109	486	1527	644	157	1193					
Eukaryotic translation	Q15056	220	24	5	886	731	689	373	342	2437	555	771	299	644					
Serotransferrin precuror	P02787	488	18	11	3113	2128	4005	548	1115	1026	1730	1408	462	2501					
galectin-1	P09382	330	31	6	1328	582	885	484	126	431	473	1157	161	110					
Tubulin beta-2 chain	P07437	148	6	2	662	352	1170	899	1198	1410	3242	394	606	2057	51	148	127	124	145
protein disulfide-isomerase A3 precursor	P30101	1210	38	21	1395	1267	1135	1724	1533	1571	2456	475	827	1447	181	391	n	257	174
ribosome-binding protein 1	Q9P2E9	540	14	18	417	816	1109	296	253	1006	373	107	116	196	n	85.5	64	174	105
Heat shock cognate 71 kDa protein	P11142	960	29	15	1740	893	n	2042	3301	9873	1216	277	248	1134		485	236	300	133
Alpha-1-antitrypsin precursor	P01009	640	35	13	304	484	2468	819	1041	1064	514	1385	458	1198		470	305	722	535
Lamin A/C.	P02545	490	27	9	379	740	467	269	229	457	339	97	105	178	n	77.5	58	157	95.1
60 kDa heat shock protein, mitochondrial precur	P10809	1110	32	20	309	422	599	625	668	557	456	220	281	3003		159	70.9	88	83.6
Heat shock 70 kDa protein 1	P08107	370	19	9	175	280	1427	2020	1018	1652	1606	800	264	692		324	255	418	310
Interferon-induced GTP-binding protein Mx1	P20591	620	23	8	173	236	334	349	373	311	255	123	157	1678		88.8	39.6	49.2	46.7
Fibrinogen gamma chain precursor	P02679	610	32	13	1153	3390	8500	434	545	2688	1804	610	1231	1418		464		975	804
ATP synthase beta chain, mitochondrial precurs	P25705	540	29	10	150	206	291	304	325	271	222	107	137	1461		77.3	34.5	42.8	40.7
Tubulin-specific chaperone A	O75347	280	54	6	319	146	182	601	293	262	157	90.7	127	228		81	50	43	106
Zinc-alpha-2-glycoprotein precursor	P25311	530	11	4	374	175	465	430	283	399	209	303	184	356	n	94.5	121	128	134
SPUF protein precursor	Q9UMX5	160	31	4	113	52.9	141	130	85.6	121	63.1	91.4	55.4	107	n	28.5	36.6	38.7	40.3
Astrocytic phosphoprotein PEA-15	Q15121	550	28	3	825	187	676	484	245	867	914	322	609	1281					
thioredoxin	P10599	110	8.6	2	275	214	190	287	166	297	142	357	83	181					
nucleobindin 1 precursor	Q02818	60	7	3	419	306	422	177	237	1165	422	245	702	533		151	80	105	131
TAR DNA-binding protein 43	Q13148	110	7	3	625	409	1396	816	732	1516	352	358	373	370		83	249	68	125
Lamin B2.	Q03252	790	27	17	456	401	226	321	288	309	170	301	556	318					
Heterogeneous nuclear ribonucleoprotein D0	Q14103	288	14	4	328	151	187	618	302	269	161	93.3	130	235		70	96	43	50
Desmin	P17661	444	32	16	252	139	337	249	170	334	212	94.7	210	271			65.3		52.7
Myosin light polypeptide 6	P60660	290	32	7	118	129	121	97.7	162	240	96.4	212	152	226					
Tropomyosin alpha 4 chain	P67936	530	23	7	456	401	226	321	288	309	170	301	556	318					
Tropomyosin 1 alpha chain	P09493	260	23	9	371	557	602	446	1209	965	555	179	471	639			90		
Tropomyosin beta chain	P07951	270	26	6	151	120	113	90.9	150	224	89.7	211	141	211					
Tumor protein D54	O43399	680	38	6	1020	232	836	599	303	1072	1131	398	753	1583					
Tumor protein D52	P55327	120	27	4	48.9	53.3	50.1	40.4	66.9	99.4	39.9	87.8	62.8	93.7					
Vimentin	P08670	580	42	15	259	644	494	333	258	733	249	160	422	635			95		
Gelsolin precursor	P06396	220	12	9	126	69.7	168	124	85	167	106	47.3	105	135			32.7		26.3
78 kDa glucose-regulated protein precursor	P11021	510	27	15	531	563	666	547	644	728	197	228	439	1047			120		

Figure 4.5 Differentially expressed proteins in stage I versus stage III/IV ovarian serous carcinoma tumors. Each cell in the matrix represents the expression level of a single protein in a single sample, with red or orange and blue or dark blue indicating intensity above and below the median for that protein across all that group samples, respectively.

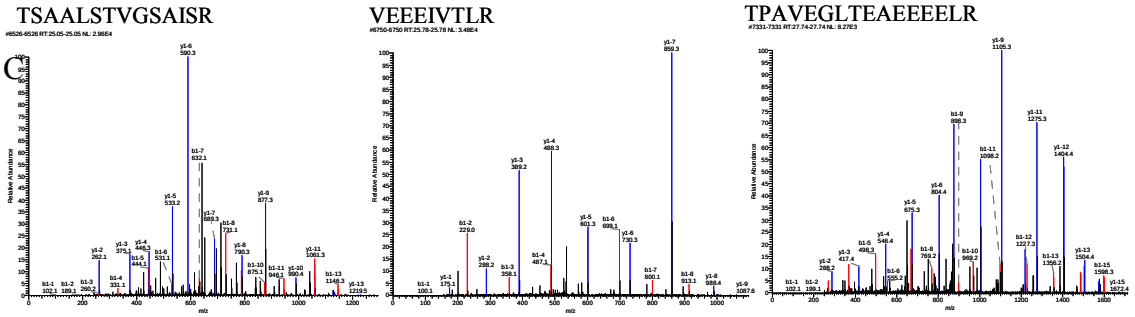
A



B



C



MDSAGQDINLNSPNK**GLLSDSMTDVPVDTGVAARTPAVEGLTEAEEEELR**AELTK**VEEEIVTLR**QVLAAKERHCGELKRR**LGLST LGELK**QNLSRSWHDVQVSSAYVKTSEKLGEWNEKV**TQSDLYKKTQETLSQAGQKTSAALSTVGSAISR**KLGDMRNSATFKSFEDR VGTIKSKVVGDRENGSDNLPSSAGSGDKPLSDPAPF

Figure 4.6 A. Protein expression maps for a given pH fraction of all tumors analyzed for Tumor protein D54. B. Protein expression levels for Tumor protein D54 among all tumors analyzed. C. LTQ analysis of several peptides identified as Tumor protein D54.

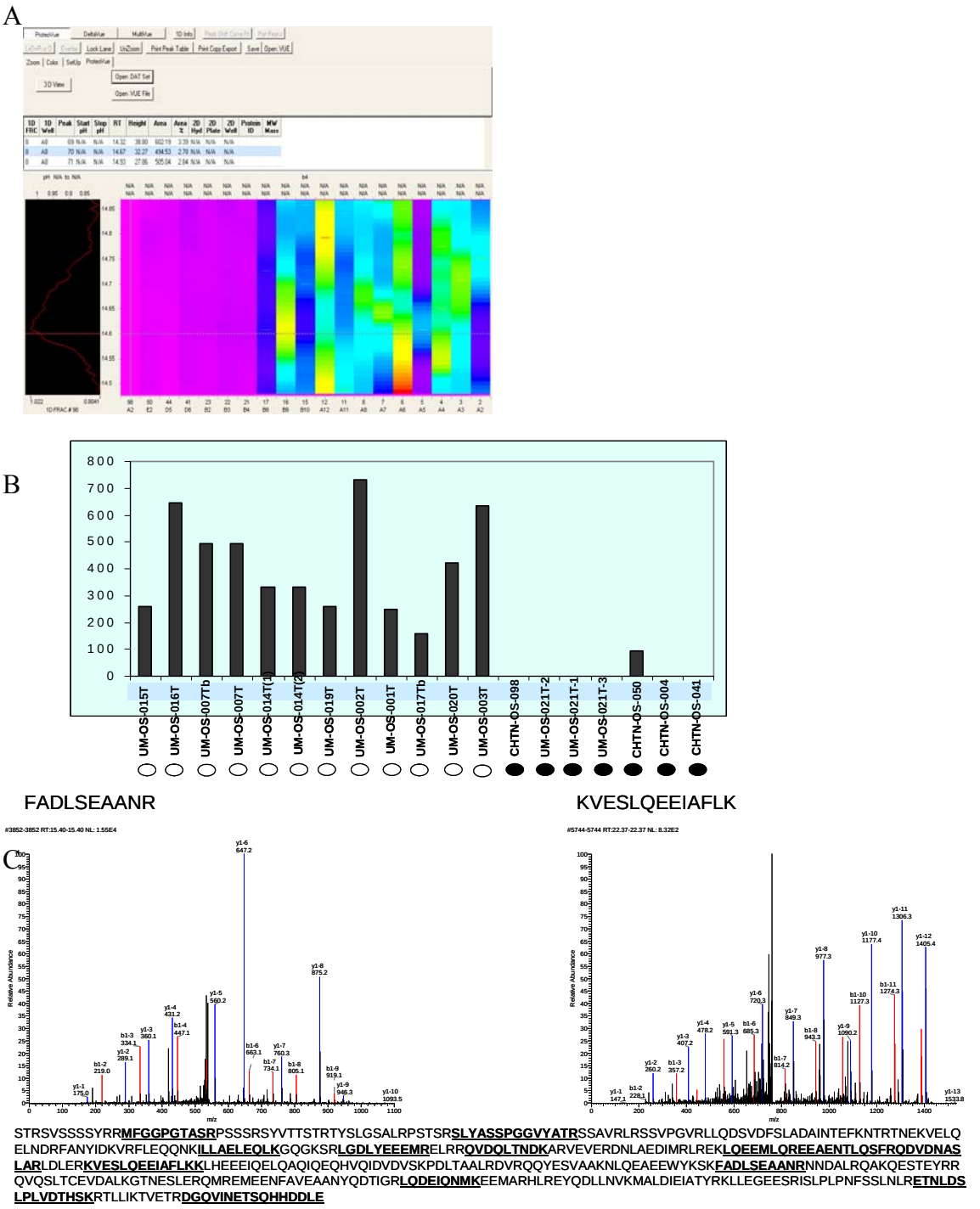
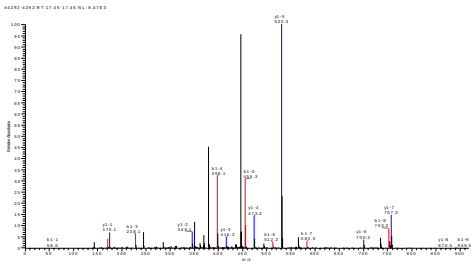
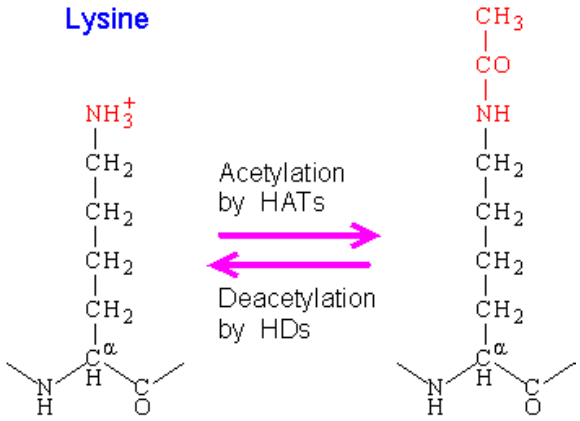


Figure 4.7 A. Protein expression maps for a given pH fraction of all tumors analyzed for Vimentin. B. Protein expression levels for Vimentin among all tumors analyzed. C. LTQ analysis of several peptides identified as Vimentin.

A K.GLGK~GGAK~R.H



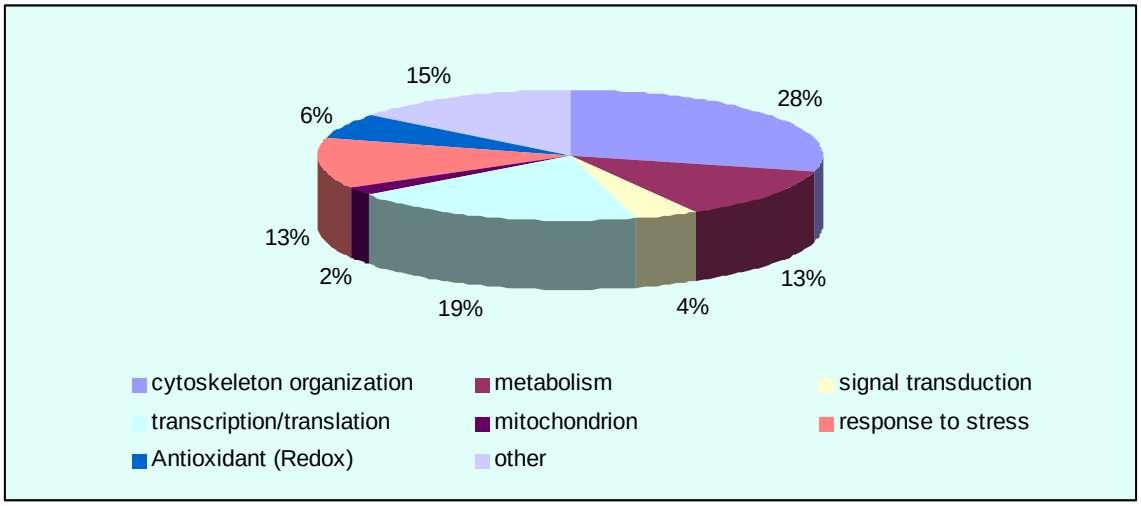
B



Acetylation and deacetylation of the lysine residue.

Figure 4.8 Histone A**.** Histone peptide with a 2 acetylation modification. B. Diagram showing putative acetylation and deacetylation of the lysine residue.

A



.

B

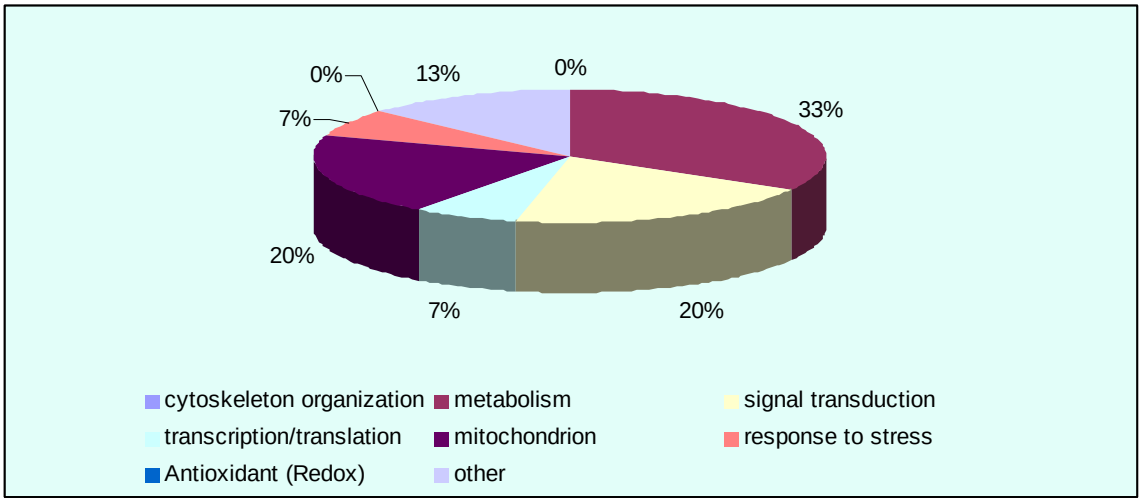


Figure 4.9 Distribution of identified proteins which are differentially expressed between low stage (stage I) and high stage (stage III/IV) tumors. **A,** over-expressed proteins in the high stage tumors **B.** Under-expressed proteins in the high stage tumors.

4.5 References:

1. Bonome, T.,Lee, J. Y.,Park, D. C.,Radonovich, M.,Pise-Masison, C.,Brady, J.,Gardner, G. J.,Hao, K.,Wong, W. H.,Barrett, J. C.,Lu, K. H.,Sood, A. K.,Gershenson, D. M.,Mok, S. C., andBirrer, M. J. (2005) Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary *Cancer Research*, *65*, 10602-10612.
2. Jemal, A.,Siegel, R.,Ward, E.,Murray, T.,Xu, J. Q.,Smigal, C., andThun, M. J. (2006) Cancer statistics, 2006 *Ca-a Cancer Journal for Clinicians*, *56*, 106-130.
3. Eltabbakh, G. H.,Yadev, P. R., andMorgan, A. (1999) Clinical picture of women with early stage ovarian cancer *Gynecologic Oncology*, *75*, 476-479.
4. Nagele, F.,Petru, E.,Medl, M.,Kainz, C.,Graf, A. H., andSevelda, P. (1995) Preoperative Ca-125 - an Independent Prognostic Factor in Patients with Stage-I Epithelial Ovarian-Cancer *Obstetrics and Gynecology*, *86*, 259-264.
5. Schwartz, D. R.,Wu, R.,Kardia, S. L. R.,Levin, A. M.,Huang, C. C.,Shedden, K. A.,Kuick, R.,Misek, D. E.,Hanash, S. M.,Taylor, J. M. G.,Reed, H.,Hendrix, N.,Zhai, Y.,Fearon, E. R., andCho, K. R. (2003) Novel candidate targets of beta-catenin/T-cell factor signaling identified by gene expression profiling of ovarian endometrioid adenocarcinomas *Cancer Research*, *63*, 2913-2922.
6. Schaner, M. E.,Ross, D. T.,Ciaravino, G.,Sorlie, T.,Troyanskaya, O.,Diehn, M.,Wang, Y. C.,Duran, G. E.,Sikic, T. L.,Caldeira, S.,Skomedal, H.,Tu, I. P.,Hernandez-Boussard, T.,Johnson, S. W.,O'Dwyer, P. J.,Fero, M. J.,Kristensen, G. B.,Borresen-Dale, A. L.,Hastie, T.,Tibshirani, R.,van de Rijn, M.,Teng, N. N.,Longacre, T. A.,Botstein, D.,Brown, P. O., andSikic, B. I. (2003) Gene expression patterns in ovarian carcinomas *Molecular Biology of the Cell*, *14*, 4376-4386.
7. Schwartz, D. R.,Kardia, S. L. R.,Shedden, K. A.,Kuick, R.,Michailidis, G.,Taylor, J. M. G.,Misek, D. E.,Wu, R.,Zhai, Y. L.,Darrah, D. M.,Reed, H.,Ellenson, L. H.,Giordano, T. J.,Fearon, E. R.,Hanash, S. M., andCho, K. R. (2002) Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas *Cancer Research*, *62*, 4722-4729.
8. Alaiya, A. A.,Franzen, B.,Hagman, A.,Dysvik, B.,Roblick, U. J.,Becker, S.,Moberger, B.,Auer, G., andLinder, S. (2002) Molecular classification of borderline ovarian tumors using hierarchical cluster analysis of protein expression profiles *International Journal of Cancer*, *98*, 895-899.
9. Hess, V.,A'Hern, R.,Nasiri, N.,King, D. M.,Blake, P. R.,Barton, D. P. J.,Shepherd, J. H.,Ind, T.,Bridges, J.,Harrington, K.,Kaye, S. B., andGore, M. E. (2004) Mucinous epithelial ovarian cancer: A separate entity requiring specific treatment *Journal of Clinical Oncology*, *22*, 1040-1044.
10. Singer, G.,Kurman, R. J.,Chang, H. W.,Cho, S. K. R., andShih, I. M. (2002) Diverse tumorigenic pathways in ovarian serous carcinoma *American Journal of Pathology*, *160*, 1223-1228.
11. Seidman, J. D., andKurman, R. J. (1996) Subclassification of serous borderline tumors of the ovary into benign and malignant types - A clinicopathologic study of 65 advanced stage cases *American Journal of Surgical Pathology*, *20*, 1331-1345.

12. Sehdev, A. E. S.,Sehdev, P. S., andKurman, R. J. (2003) Noninvasive and invasive micropapillary (low-grade) serous carcinoma of the ovary - A clinicopathologic analysis of 135 cases *American Journal of Surgical Pathology*, *27*, 725-736.
13. Silva, E. G.,Tornos, C. S.,Malpica, A., andGershenson, D. M. (1997) Ovarian serous neoplasms of low malignant potential associated with focal areas of serous carcinoma *Modern Pathology*, *10*, 663-667.
14. Shih, I. M., andKurman, R. J. (2004) Ovarian tumorigenesis - A proposed model based on morphological and molecular genetic analysis *American Journal of Pathology*, *164*, 1511-1518.
15. Ahmed, N.,Oliva, K. T.,Barker, G.,Hoffmann, P.,Reeve, S.,Smith, I. A.,Quinn, M. A., andRice, G. E. (2005) Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer *Proteomics*, *5*, 4625-4636.
16. Jones, M. B.,Krutzsch, H.,Shu, H. J.,Zhao, Y. M.,Liotta, L. A.,Kohn, E. C., andPetricoin, E. F. (2002) Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer *Proteomics*, *2*, 76-84.
17. Petricoin, E. F.,Ardekani, A. M.,Hitt, B. A.,Levine, P. J.,Fusaro, V. A.,Steinberg, S. M.,Mills, G. B.,Simone, C.,Fishman, D. A.,Kohn, E. C., andLiotta, L. A. (2002) Use of proteomic patterns in serum to identify ovarian cancer *Lancet*, *359*, 572-577.
18. Kachman, M. T.,Wang, H. X.,Schwartz, D. R.,Cho, K. R., andLubman, D. M. (2002) A 2-D liquid separations/mass mapping method for interlysate comparison of ovarian cancers *Analytical Chemistry*, *74*, 1779-1791.
19. Wang, H. X.,Kachman, M. T.,Schwartz, D. R.,Cho, K. R., andLubman, D. M. (2004) Comprehensive proteome analysis of ovarian cancers using liquid phase separation, mass mapping and tandem mass spectrometry: A strategy for identification of candidate cancer biomarkers *Proteomics*, *4*, 2476-2495.
20. Hamler, R. L.,Zhu, K.,Buchanani, N. S.,Kreunin, P.,Kachman, M. T.,Miller, F. R., andLubman, D. M. (2004) A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification *Proteomics*, *4*, 562-577.
21. Russell, P. In *Blausteins pathology of the femal genital tract*; SpringerVerlag: New York, 1994, pp 705-782.
22. Scully, R. E.,Yound, R. H., andClement, P. B. *Tumors of the Ovary, Maldeveloped Gonads, Fallopian Tube, and Broad Ligament*; Armed Forces Institute of Pathology: washington, D. C., 1998.
23. Wang, Y. F.,Wu, R.,Cho, K. R.,Shedden, K. A.,Barder, T. J., andLubman, D. M. (2006) Classification of cancer cell lines using an automated two-dimensional liquid mapping method with hierarchical clustering techniques *Molecular & Cellular Proteomics*, *5*, 43-52.
24. Liu, H. B.,Sadygov, R. G., andYates, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics *Analytical Chemistry*, *76*, 4193-4201.
25. Byrne, J. A.,Tomasetto, C.,Garnier, J. M.,Rouyer, N.,Mattei, M. G.,Bellocq, J. P.,Rio, M. C., andBasset, P. (1995) A Screening Method to Identify Genes

Commonly Overexpressed in Carcinomas and the Identification of a Novel Complementary-DNA Sequence *Cancer Research*, *55*, 2896-2903.

26. Byrne, J. A.,Balleine, R. L.,Fejzo, M. S.,Mercieca, J.,Chiew, Y. E.,Livnat, Y.,St Heaps, L.,Peters, G. B.,Byth, K.,Karlan, B. Y.,Slamon, D. J.,Harnett, P., andDefazio, A. (2005) Tumor protein D52 (TPD52) is overexpressed and a gene amplification target in ovarian cancer *International Journal of Cancer*, *117*, 1049-1054.
27. Rubin, M. A.,Varambally, S.,Beroukhim, R.,Tomlins, S. A.,Rhodes, D. R.,Paris, P. L.,Hofer, M. D.,Storz-Schweizer, M.,Kuefer, R.,Fletcher, J. A.,Hsi, B. L.,Byrne, J. A.,Pienta, K. J.,Collins, C.,Sellers, W. R., andChinnaiyan, A. M. (2004) Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer *Cancer Research*, *64*, 3814-3822.
28. Wang, R. X.,Xu, J. C.,Saramaki, O.,Visakorpi, T.,Sutherland, W. M.,Zhou, J. G.,Sen, B.,Lim, S. D.,Mabjeesh, N.,Amin, M.,Dong, J. T.,Petros, J. A.,Nelson, P. S.,Marshall, F. F.,Zhau, H. E., andChung, L. W. K. (2004) PrLZ, a novel prostate-specific and androgen-responsive gene of the TPD52 family, amplified in chromosome 8q21.1 and overexpressed in human prostate cancer *Cancer Research*, *64*, 1589-1594.
29. Balleine, R. L.,Fejzo, M. S.,Sathasivam, P.,Basset, P.,Clarke, C. L., andByrne, J. A. (2000) The hD52 (TPD52) gene is a candidate target gene for events resulting in increased 8q21 copy number in human breast carcinoma *Genes Chromosomes & Cancer*, *29*, 48-57.
30. Boutros, R.,Fanayan, S.,Shehata, M., andByrne, J. A. (2004) The tumor protein D52 family: many pieces, many puzzles *Biochemical and Biophysical Research Communications*, *325*, 1115-1121.
31. Fraga, C. H.,True, L. D., andKirk, D. (1998) Enhanced expression of the mesenchymal marker, vimentin, in hyperplastic versus normal human prostatic epithelium *Journal of Urology*, *159*, 270-274.
32. Cuezva, J. M.,Krajewska, M.,de Heredia, M. L.,Krajewski, S.,Santamaria, G.,Kim, H.,Zapata, J. M.,Marusawa, H.,Chamorro, M., andReed, J. C. (2002) The bioenergetic signature of cancer: A marker of tumor progression *Cancer Research*, *62*, 6674-6681.

Chapter 5

Lectin Affinity as an Approach to the Proteomic Analysis of Membrane Glycoproteins for Breast Cancer Cell Lines

5.1 Introduction

Glycoproteins are proteins which have the attachment of specific carbohydrate structures which play a major role in determining protein function. Glycosylation alterations have been associated with the development and progression of cancer. A large number of clinical biomarkers and therapeutic targets in cancer are glycoproteins (1-5), such as CA125 in ovarian cancer, Her2/neu in breast cancer, and prostate-specific antigen (PSA) in prostate cancer. Monitoring glycoproteins could play an important role in detecting and evaluating tumor progression in addition to assessing tumor load and therapy.

Lectins are proteins that specifically detect carbohydrate components and its specificity has provided the basis for probing different types of glycoconjugates in the cellular context. Lectin affinity chromatography has recently been widely used to purify glycoproteins (6-8) with specific structures and to examine their microheterogeneity to the level of a single sugar residue (9).

Membrane proteins play critical roles in many biological functions and are often the molecular targets for drug discovery. Membrane proteins are often extensively glycosylated, especially those that decorate the extracellular membrane (4,7). Affinity

purification based on the properties of the glycoforms attached to these proteins has been used extensively to enrich membrane proteins (4,7).

In this work, we demonstrated the use of combining membrane extraction and lectin affinity column as a simple method to apply the proteomic analysis of membrane glycoproteins between two breast cancer cell lines CA1a and AT1. We first compared two different lysis methods for the membrane proteins extraction, then we applied lectin affinity extraction using ConA to extract the glycoproteome of these two cell lines. The protein expression between these two different cell lines was compared separately. A spectral count label-free method was applied for the quantitation of isolated membrane glycoproteins between these two cell lines. We could find differentially expressed proteins between these two different cell lines based on their glycoproteome expression to identify potential markers using this methodology.

5.2 Experimental Section

5.2.1 Cell Culture

Fully malignant human breast cancer cells, MCF10CA1a cells and MCF10AT1 cells were grown in monolayer on plastic in DMEM/F12 medium (1:1 mixture of DMEM and Ham's F-12 medium) supplemented with 5% horse serum, 10 mg/mL of insulin, 20 ng/mL epidermal growth factor, and 0.5 mg/mL of hydrocortisone. The cells were then collected by scraping and washed twice with PBS buffer before being stored at -80ºC.

5.2.2 Lysis

5.2.2.1 Lysis method 1

The packed cells, which is around 0.2ml were suspended in 1.5ml of lysis buffer containing 10mM Tris-HCl, pH-7.5, 150 mM NaCl, 1% Nonidet P-40 and protease

inhibitor cocktail (Roche) for 30 min and use a tissue tearor (Biospec Products, Inc) for a total of 1 min in 15 s pulses with 30 s cooling periods after each pulse, the speed can reach to the 30,000 RPM. The lysate was centrifuged at 132 000 rpm for 30 min at 4°C, and the supernatant was collected and stored at -80 °C if not used immediately. Cell lysates were prefiltered, 0.45 *i*m filter (Millipore), prior to loading them on to columns. The final samples were subjected to quantitation by Bradford assay (Bio-Rad, Hercules, CA).

5.2.2.2 Lysis method 2 – MEM-PER Eukaryotic membrane protein extraction reagent kit

Totally300 µl of Reagent A were added to the cell pellet. We pipetted up and down to obtain a homogeneous cell suspension and then incubated for 10 minutes at room temperature with occasional vortexing. Lysed cells were place on ice. Two parts Reagent C was diluted with 1 part Reagent B which made sufficient mixture for each sample to receive 900 µl. The 900 µl diluted reagent C were added to each tube of lysed cells and vortexed. After that, tubes were incubated on ice for 30 minutes, vortexed every 5 minutes. We centrifuged tubes at 10,000 x *g* for 3 minutes at 4°C and transferred supernatant to new tubes. Supernatant were incubated for 20 minutes at 37°C to separate the membrane protein fraction. Tubes were then centrifuge at room temperature for 2 minutes at 10,000 x *g* to isolate the hydrophobic fraction from the hydrophilic fraction. Finally we carefully removed the hydrophilic phase (top layer) from the hydrophobic protein phase (bottom layer) and saved in a new tube. Perform the phase separations as quickly as possible because the interface between the layers slowly disappears at room

temperature. The final samples were subjected to quantitation by Bradford assay (Bio-Rad, Hercules, CA). The samples were put into -80ºC freezer if not used.

5.2.3 Lectin Affinity Glycoprotein Extraction.

Agarose-bound Concanavalin A (ConA) were purchased from Vector Laboratories (Burlingame, CA). Agarose-bound 1 mL ConA were packed into disposable screw endcap spin columns with filters at both ends. The column was first washed with 3 mL of binding buffer (20 mM Tris, 0.15 M NaCl, pH 7.4). Protease inhibitor stock solution was prepared by dissolving one complete EDTA-free protease inhibitor cocktail tablet (Roche, Indianapolis, IN) in 1 mL of H_2O. The stock solution was added to the binding buffer and elution buffer at a ratio of (v/v) 1:50. Before running the lectin column, the lysate was diluted with 10 times binding buffer and concentrated to around 700ul using Microcon YM-10 (Millipore Corp., Bradford, MA) according to the manufacturer's protocol. A 500µL sample diluted with 5mL of binding buffer was loaded onto the column and incubated for 15 min. The column was washed with 6 mL of binding buffer twice to wash off the nonspecific binding. The proteins were resuspended in binding buffer and then passed through the lectin affinity column. The captured glycoproteins were released with 10mL of elution buffer (0.3 M methyl-R-D-mannopyroside in 20 mM Tris and 0.5 M NaCl, pH 7.0). This step was repeated twice and the eluted fractions were pooled. The sample was concentrated using Microcon YM-10 (Millipore Corp., Bradford, MA) according to the manufacturer's protocol to 400ul. The final eluted samples were subjected to quantitation by Bradford assay (Bio-Rad, Hercules, CA).

5.2.4 Protein Digestion by Trypsin

Ten microliters of 50 mM ammonium bicarbonate (Sigma) was mixed with each concentrated 10 microliters sample to obtain pH 7.8. A 2 µL portion of 100 mM DTT (Sigma) was added and the resulting mixture incubated at 60 °C for 60 min. Another 180 µL ammonium bicarbonate was added. TPCK modified sequencing grade porcine trypsin (1:50) (Promega, Madison, WI) was added and briefly vortexed prior to a 12-16 h incubation at 37 °C on an agitator. The digestion and reduction reaction was terminated by adding 1 µL of TFA to the digest.

5.**2.5 Mass Spectrometry. Protein Identification by LC-MS/MS.**

A Paradigm MG4 micropump (Michrom Biosciences Inc., Auburn, CA) was used for chromatographic separation of peptide mixtures. For nanoLC-ESI-MS/MS experiments, a nanotrap platform (Michrom) was set up prior to the electrospray source. It included a peptide nanotrap (0.2 × 50 mm^2, Michrom) and a separation column (0.1 mm × 150 mm, C18, Michrom). The peptide sample was injected and first desalted on the trap column with 3% solvent B (0.3% formic acid in 98% ACN) at 50 µL/min for 5 min. The peptides were then eluted using a 45 min gradient from 5 to 95% B at a flow rate of 0.3 µL/min where solvent A was 0.3% formic acid in HPLC grade water. A Finnigan LTQ mass spectrometer (Thermo) was used to acquire spectra. A 75 µm metal spray tip (Michrom) was used, and spray voltage was set at 2.5 kV. The instrument was operated in data-dependent mode with dynamic exclusion enabled. The MS/MS spectra on the five most abundant peptide ions in full MS scan were obtained. The data acquisition and generation of peak list files were automatically done by Xcaliber software.

5.2.6 Database Searching and Quantitation by Lable-free Spectral Count.

All MS/MS spectra were searched against the human protein database from SwissProt using SEQUEST algorithm incorporated in Bioworks software, version 3.1 SR1 (Thermo Finnigan). Peptide fragment lists were generated and submitted to Swiss-Prot database searching. The search parameters were as follows: (1) database species, *Homo sapiens*; (2) allowing two missed cleavages; (3) possible modifications, oxidation of M; (4) peptide ion mass tolerance 1.50 Da; (5) fragment ion mass tolerance 0.0 Da; (6) peptide charges +1, +2, and +3. The filter function in Bioworks browser was used to set a single threshold to consider fully tryptic peptides assigned with Xcorr values as follows: 1.9 for singly charged ions, 2.5 for doubly charged ions, and 3.5 for triply charged ions, while no ions at higher charged states were considered. ΔCn cutoff was set as 0.1. All SEQUEST search parameters and data filtering were the same in all digest fractions. Protein IDs were accepted if, and only if, the ID was positively identified in at least two MS/MS analyses. If the protein was identified by a single peptide matching, the spectrum was manually validated. The matched ions covering at least 70% of the peptide sequence must have a high signal-to- noise ratio (S/N). Two high S/N spectra of a single peptide matching protein were accepted for positive identification. Positive protein identification was validated by the Trans- Proteomics pipeline. This software includes both the PeptideProphet and ProteinProphet programs that were developed by Keller et al. (http://peptideprophet.sourceforge.net/). All the reported proteins have an identification probability higher than 90% and protein and peptide confidence above 90%.

The spectral count is a mass spectrometry (MS)-based *label-free quantitation* approach. It counts the total number of MS/MS spectra taken on peptides from a given protein in a given LC/LC-MS/MS analysis. A number of studies have demonstrated a

high correlation between spectral count and protein abundances with the R square value around 0.99 (10) and is linearly correlated with the protein abundance over a dynamic range of 2 orders of magnitude. In addition, among three sampling statistics that are correlated with protein abundance (i.e spectral count, peptide count, sequence coverage) (11), spectral count method has the highest reproducibility and is the best performer in protein relative quantitation. We performed a comparative analysis between two samples using spectral count. Spectral count of individual proteins in both samples is acquired using DTASelect v2.0. All proteins with total spectral count less than 10 are considered absent and thus eliminated and confident proteins IDs are confirmed with the Trans- Proteomics pipeline analysis. The fold change based on spectral count is calculated followed by a t statistical test to examine the significant difference in protein spectral count between the two samples. All statistics calculation is performed in R which is a free software environment for statistical computing and graphics. A p-value of 0.05 is used to determine whether a protein is differentially expressed or not.

5.3 Results and Discussion

5.3.1 Cell lines

In this study, two cell lines were used. CA1a is a fully malignant human breast cancer cell line which rapidly forms invasive carcinomas with metastatic potential and displays histologic variations ranging from undifferentiated carcinomas to well differentiated adenocarcinomas (12). AT1 is a preneoplastic human breast cell line and this cell line forms preneoplastic lesions in nude mice that represent a premalignant stage with potential for neoplastic progression (13).

5.3.2 Comparison of Two Lysis Methods

In the present study, we investigated and compared the membrane glycoprotein extracted from two different lysis methods. In these experiments, around 3×10^7 AT1 and 10^8 CA1a cells were used. Total cell lysates of CA1a and AT1 cells were passed through a Con A lectin column to capture oligomannosyl saccharides found in N-glycans (14,15). After ConA lectin extraction, the Bradford assay method was used for the quantitation of the membrane glycoproteins. The data are shown in Table 1. The proteins after lectin extraction correspond to a 150-200 fold average enrichment in the corresponding affinity purified proteins relative to the starting materials. In our studies, we found that the lysis method 1 provides improved recovery compared to the second lysis method. We also applied AT1 samples twice as AT1-1 and AT1-2 for the reproduciblility studies. Based on our results, we observe that the lysis method and lectin extraction are very reproducible.

5.3.3 Protein Identification.

Con A bound proteins were used prior mass spectrometric analysis. The lectin extracted proteins were digested by trypsin and analyzed by nano-LC/MS/MS. For each scan, the five most abundant peptides were sequenced. Figure 1A is a representative nano-LC/MS/MS base peak chromatogram, showing the detection of the more abundant ions across a 45 min gradient separation. Figure 1B shows a representative MS/MS spectrum of a peptide sequence from endoplasmin precursor, one of the N-linked glycoproteins as annotated in the Swiss-Prot database. Eighteen other ion peaks were also detected and assigned to this protein.

Tryptic digest fractions were analyzed. Positive protein identification was validated by the Trans- Proteomics pipeline. This software includes both the PeptideProphet and ProteinProphet programs that were developed by Keller et al. (http://peptideprophet.sourceforge.net/). All the reported proteins have an identification probability higher than 90% and 90% minimum peptide and protein probability. Common glycoproteins are listed in Table 2. Based on the data we obtained, they are all N-linked glycoproteins as annotated in the Swiss-Prot database. In terms of the subcellular location, around 50% are in the endoplasmic reticulum lumen.

5.3.4 Analysis of Isolated Proteins from ConA for CA1a and AT1

With the ConA lectin extraction, there was a total of 153 unique proteins and 515 unique peptides were identified in CA1a and 104 unique proteins and 326 unique peptides in AT1. The distributions of the molecular weights and pI range of the isolated proteins are shown in Figure 2. For AT1 in Figure 2A, the molecular weights ranged from ~10 to ~180 kDa, which is very similar to the range with CA1a in Figure 2C. Similarly analysis of the pI values of these proteins in Figure 2B for AT1 and 2D for CA1a indicated the capacity to capture proteins with a wide range of values ranged from pH 4 to pH12. There did not appear to be any major differences in the molecular weight or pI ranges of the proteins selected by ConA.

5.**3.5 Differentially Expressed Glycoproteins Analysis.**

We performed a comparative analysis between two samples using spectral count. Spectral count of individual proteins in both samples is acquired using DTASelect v2.0. All proteins with a total spectral count less than 10 are considered absent and thus eliminated and confident proteins IDs are confirmed with the Trans- Proteomics pipeline

analysis. The fold change based on spectral count is calculated followed by a t statistical test to examine the significant difference in protein spectral count between the two samples. All statistics calculations are performed in R. A p-value of 0.05 is used to determine whether a protein is differentially expressed. The differentially expressed glycoproteins with their fold change are shown in Table 3. The ratio here is a normalized ratio, that takes into account the variation between the total spectral count between the two samples, and it also adds 1.25 (correction factor) to every spectral count (such that no zero value occurs). The specific formula for the normalized ratio is:

Rsc = log2[(n2+f)/(n1+f)] + log2[(t1-n1+f)/(t2-n2+f)]

Highly differentially expressed proteins were further evaluated as possible candidate biomarkers to distinguish between CA1a and AT1. We focused on several proteins previously associated with cancer that displayed the largest-fold change. Gamma-glutamyl hydrolase is a lysosomal enzyme involved in the metabolism of folates and anti-folates and it plays a role as a tumor marker in breast and lung cancer (18). In our studies, we found that this protein only appears in CA1a and not in AT1 which is consistant with previous studies (19,20). The LC/MS/MS spectrum is shown in Figure 3.

5.4 Conclusion

In this work, a fully malignant human breast cancer cell line CA1a and a preneoplastic human breast cell line AT1 were studied. We demonstrated the use of lectin affinity column as an approach for proteomic analysis of membrane glycoproteins. In our studies, we found that lysis method 1 is better than the membrane extract kit for our strategies. After lectin extraction, it was shown that there was selective enrichment of proteins by the lectin affinity steps and both cell line samples have their own specific

glycoprotein expression patterns. However, there are few major differences in the pI or molecular weight ranges of the proteins selected by ConA for both the CA1a and AT1 cell lines. Spectral count lable-free method was found to be a useful approach for the quantitation of isolated membrane glycoproteins between these two cell lines. Differentially expressed proteins identified between these two different cell lines could facilitate biomarker discovery. The use of lectin columns in conjunction with mass spectrometry offers a useful approach for the isolation and identification of novel glycoproteins. This work expands our tissue proteome capabilities from the analysis of soluble proteins in previous studies to the examination of membrane proteins.

Method1	Total cells	After Lectin(ug)	Method2	Total cells	After Lectin(ug)
AT1	3×10^7	15.9	AT1	3×10^7	11.8
AT1-2	3×10^7	16.8	AT1-2	3×10^7	10.5
CA1a	10^8	126.9	CA1a	10^8	87.0

Table 5.1 Comparison between two different lysis methods.

Swiss-Prot ID	Protein Name	Subcellular location	Glycoprotein
SUMF2_HUMAN	Sulfatase-modifying factor 2 precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
BASI_HUMAN	Basigin precursor	Cell membrane	N-linked (GlcNAc...)
RCN1_HUMAN	Reticulocalbin-1 precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
CATD_HUMAN	Cathepsin D precursor	Lysosome.	N-linked (GlcNAc...)
4F2_HUMAN	4F2 cell-surface antigen heavy chain	Membrane	N-linked (GlcNAc...)
ENPL_HUMAN	Endoplasmin precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
HEXA_HUMAN	Beta-hexosaminidase alpha chain precursor	Lysosome	N-linked (GlcNAc...)
ASPH_HUMAN	Aspartyl\asparaginyl beta-hydroxylase	Endoplasmic reticulum membrane	N-linked (GlcNAc...)
OXRP_HUMAN	150 kDa oxygen-regulated protein precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
DSG2_HUMAN	Desmoglein-2 precursor	Cell membrane	N-linked (GlcNAc...)
BGAL_HUMAN	Beta-galactosidase precursor	Lysosome	N-linked (GlcNAc...)
UGGG1_HUMAN	UDP-glucose:glycoprotein glucosyltransferase 1 precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
CALU_HUMAN	Calumenin precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
HEXB_HUMAN	Beta-hexosaminidase beta chain precursor	Lysosome	N-linked (GlcNAc...)
GLU2B_HUMAN	Glucosidase II beta subunit precursor	Endoplasmic reticulum	N-linked (GlcNAc...)
CADH1_HUMAN	Epithelial-cadherin precursor	Cell membrane	N-linked (GlcNAc...)
GANAB_HUMAN	Neutral alpha-glucosidase AB precursor	Endoplasmic reticulum	N-linked (GlcNAc...)
SPH2_HUMAN	Collagen-binding protein 2 precursor	Endoplasmic reticulum lumen	N-linked (GlcNAc...)
FUCO2_HUMAN	Plasma alpha-L-fucosidase precursor	Secreted	N-linked (GlcNAc...)
KE4_HUMAN	Zinc transporter SLC39A7	Membrane	N-linked (GlcNAc...)

Table 5.2 Common glycoproteins ID in CA1a and AT1.

Swiss-Prot ID	Protein Name	AT1	CA1a	Rsc Ratio	P value
RCN1_HUMAN	Reticulocalbin-1 precursor	262	39	-3.80	5.55E-16
CATD_HUMAN	Cathepsin D precursor	878	92	-4.58	0
4F2_HUMAN	4F2 cell-surface antigen heavy chain	36	0	-5.92	2.04E-6
ENPL_HUMAN	Endoplasmin precursor	2004	369	-4.36	0
HEXA_HUMAN	Beta-hexosaminidase alpha chain precursor	10	0	-4.18	0.045
ASPH_HUMAN	Aspartyl\asparaginyl beta-hydroxylase	123	20	-3.59	2.4E-7
OXRP_HUMAN	150 kDa oxygen-regulated protein precursor	245	0	-8.72	0
DSG2_HUMAN	Desmoglein-2 precursor	20.2	0	-7.23	0
CALU_HUMAN	Calumenin precursor	58	29	-1.99	0.925
HEXB_HUMAN	Beta-hexosaminidase beta chain precursor	55	0	-6.52	1.11E-9
GLU2B_HUMAN	Glucosidase II beta subunit precursor	0	20	3.09	7.11E-10
CADH1_HUMAN	Epithelial-cadherin precursor	15	0	-4.71	0.0069
GGH_HUMAN	Gamma-glutamyl hydrolase	4	10	-2.49	0.0008
GANAB_HUMAN	Neutral alpha-glucosidase AB precursor	483	70	-3.94	0

Table 5.3 Differentially expressed glycoproteins in CA1a and AT1.

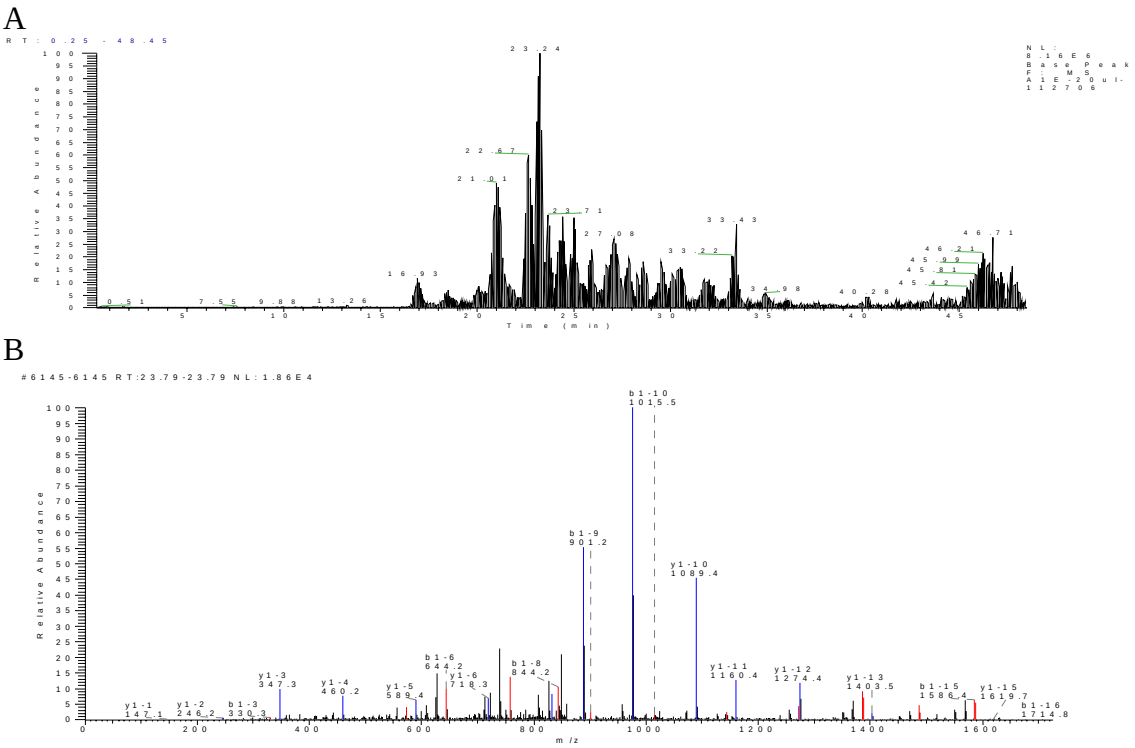


Figure 5.1 LC/MS/MS analysis. (A) A representative nano-LC/MS/MS base peak chromatogram, showing the detection of the peptide ions across the 45 min gradient separation. (B) MS/MS sequencing data of a peptide from Endoplasmin precursor identified in the eluted fraction of CA1a and AT1 cell lines

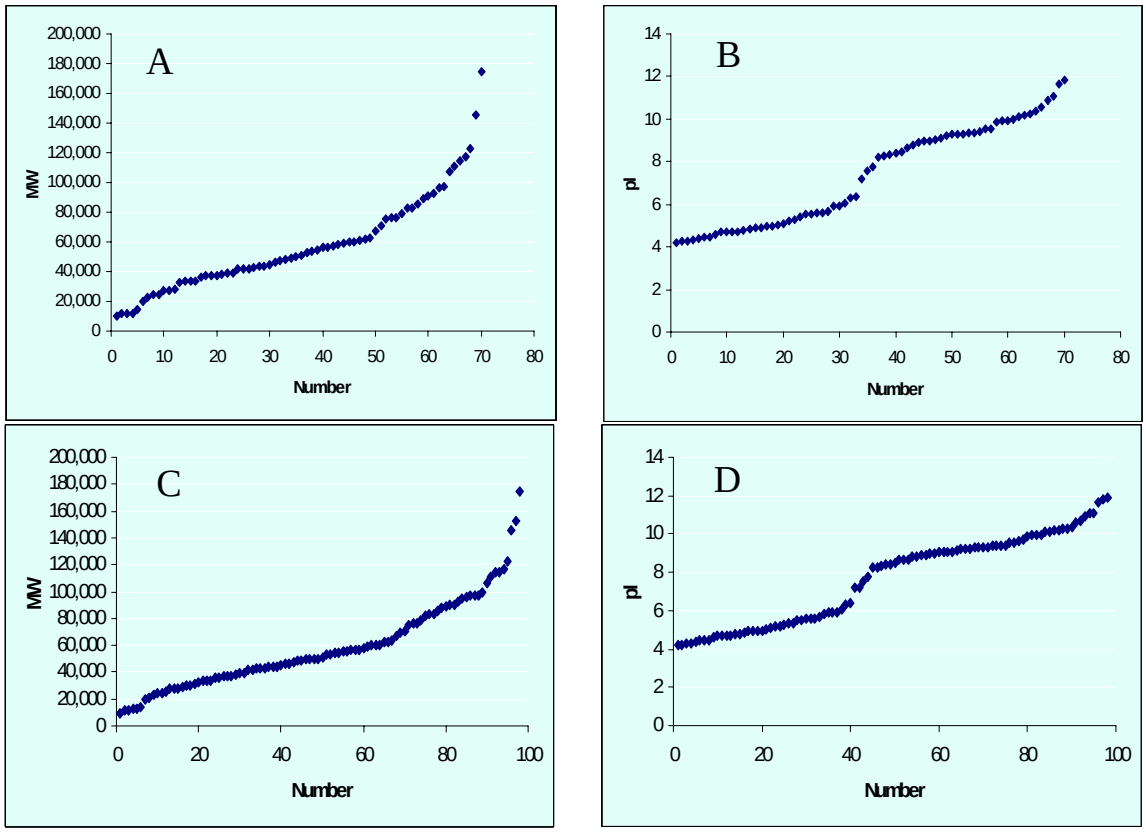


Figure 5.2 Distribution of detected molecular weights and pI values of proteins identified in the AT1 (A,B) and CA1a (C,D) using ConA

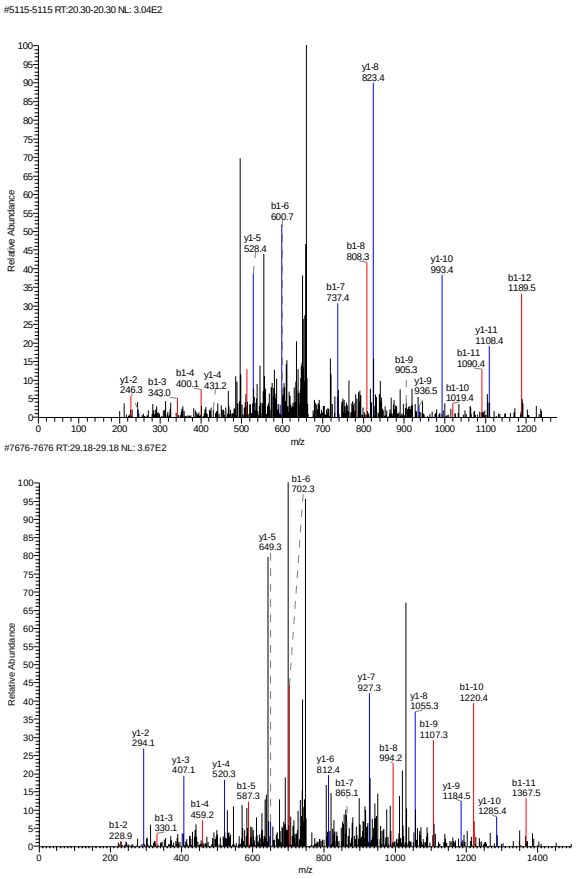


Figure 5.3 Gamma-glutamyl hydrolase

5.5 References:

1. Nakamori, S.,Ota, D. M.,Cleary, K. R.,Shirotani, K., and Irimura, T. (1994) Muc1 Mucin Expression as a Marker of Progression and Metastasis of Human Colorectal-Carcinoma *Gastroenterology, 106*, 353-361.
2. Hakkinen, I.,Nevalainen, T.,Paasivuo, R.,Partanen, P.,Seppala, K., and Sipponen, P. (1991) Gastric-Cancer Associated Structure in Mucus Glycoproteins Shown as a Clinically Useful Marker *Gut, 32*, 1465-1469.
3. Wada, I.,Rindress, D.,Cameron, P. H.,Ou, W. J.,Doherty, J. J.,Louvard, D.,Bell, A. W.,Dignard, D.,Thomas, D. Y., and Bergeron, J. J. M. (1991) Ssr-Alpha and Associated Calnexin Are Major Calcium-Binding Proteins of the Endoplasmic-Reticulum Membrane *Journal of Biological Chemistry*, *266*, 19599-19610.
4. Metzelaar, M. J.,Wijngaard, P. L. J.,Peters, P. J.,Sixma, J. J.,Nieuwenhuis, H. K., and Clevers, H. C. (1991) Cd63-Antigen - a Novel Lysosomal Membrane Glycoprotein, Cloned by a Screening-Procedure for Intracellular Antigens in Eukaryotic Cells *Journal of Biological Chemistry*, *266*, 3239-3245.
5. Fernandes, B.,Sagman, U.,Auger, M.,Demetrio, M., and Dennis, J. W. (1991) Beta-1-6 Branched Oligosaccharides as a Marker of Tumor Progression in Human Breast and Colon Neoplasia *Cancer Research*, *51*, 718-723.
6. Kawakami, E.,Hirano, T.,Hori, T., and Tsutsui, T. (2004) Protease-induced hyperactivation of canine spermatozoa associated with disappearance of lectin-binding glycoproteins on their surface *Journal of Veterinary Medical Science*, *66*, 1027-1031.
7. Lotan, R., and Nicolson, G. L. (1979) Purification of Cell-Membrane Glycoproteins by Lectin Affinity-Chromatography *Biochimica Et Biophysica Acta*, *559*, 329-376.
8. Lotan, R.,Beattie, G.,Hubbell, W., and Nicolson, G. L. (1977) Activities of Lectins and Their Immobilized Derivatives in Detergent Solutions - Implications on Use of Lectin Affinity Chromatography for Purification of Membrane Glycoproteins *Biochemistry*, *16*, 1787-1794.
9. Satish, P. R., and Surolia, A. (2001) Exploiting lectin affinity chromatography in clinical diagnosis *Journal of Biochemical and Biophysical Methods*, *49*, 625-640.
10. Liu, H. B.,Sadygov, R. G., and Yates, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics *Analytical Chemistry*, *76*, 4193-4201.
11. Zhang, B.,VerBerkmoes, N. C.,Langston, M. A.,Uberbacher, E.,Hettich, R. L., and Samatova, N. F. (2006) Detecting differential and correlated protein expression in label-free shotgun proteomics *Journal of Proteome Research*, *5*, 2909-2918.
12. Santner, S. J.,Dawson, P. J.,Tait, L.,Soule, H. D.,Eliason, J.,Mohamed, A. N.,Wolman, S. R.,Heppner, G. H., and Miller, F. R. (2001) Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells *Breast Cancer Research and Treatment*, *65*, 101-110.
13. Dawson, P. J.,Wolman, S. R.,Tait, L.,Heppner, G. H., and Miller, F. R. (1996) MCF10AT: A model for the evolution of cancer from proliferative breast disease *American Journal of Pathology, 148*, 313-319.

14. Mega, T.,Oku, H., and Hase, S. (1992) Characterization of Carbohydrate-Binding Specificity of Concanavalin-a by Competitive-Binding of Pyridylamino Sugar Chains *Journal of Biochemistry*, *111*, 396-400.
15. Ohyama, Y.,Kasai, K. I.,Nomoto, H., and Inoue, Y. (1985) Frontal Affinity-Chromatography of Ovalbumin Glycoasparagines on a Concanavalin a-Sepharose Column - a Quantitative Study of the Binding-Specificity of the Lectin *Journal of Biological Chemistry*, *260*, 6882-6887.
16. Bakry, N.,Kamata, Y., and Simpson, L. L. (1991) Lectins from Triticum-Vulgaris and Limax-Flavus Are Universal Antagonists of Botulinum Neurotoxin and Tetanus Toxin *Journal of Pharmacology and Experimental Therapeutics*, *258*, 830-836.
17. Wang, Y. H.,Wu, S. L., and Hancock, W. S. (2006) Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry *Glycobiology*, *16*, 514-523.
18. Schneider, E., and Ryan, T. J. (2006) Garnma-glutarnyl hydrolase and drug resistance *Clinica Chimica Acta*, *374*, 25-32.
19. Finak, G.,Sadekova, S.,Pepin, F.,Hallett, M.,Meterissian, S.,Halwani, F.,Khetani, K.,Souleimanova, M.,Zabolotny, B.,Omeroglu, A., and Park, M. (2006) Gene expression signatures of morphologically normal breast tissue identify basal-like tumors *Breast Cancer Research*, *8*.
20. He, P.,Varticovski, L.,Bowman, E. D.,Fukuoka, J.,Welsh, J. A.,Miura, K.,Jen, J.,Gabrielson, E.,Brambilla, E.,Travis, W. D., and Harris, C. C. (2004) Identification of carboxypeptidase E and gamma-glutamyl hydrolase as biomarkers for pulmonary neuroendocrine tumors by cDNA microarray *Human Pathology*, *35*, 1196-1209.

Chapter 6

The Study of Humoral Response in Pancreatic Cancer Using modified Protein Microarrays from Panc-1 Cell-lysates

6.1 Introduction

Cancer is a class of diseases which is characterized by uncontrolled cell division and the ability of these cells to invade other tissues, either by direct growth into adjacent tissue (invasion) or by migration of cells to distant sites (metastasis). A total of 1,399,790 new cancer cases and 564,830 deaths from cancer are expected in the United States in 2006 (1). Among them, approximately 30,000 people die of pancreatic cancer each year. Pancreatic cancer is the third most common malignancy and the fifth leading cause of cancer-related mortality. The main reason for failing to achieve a cure in most patients is because the diagnosis of pancreatic cancer is rarely made at an early stage and most patients have incurable disease by the time they present with symptoms. The overall 5-year survival rate for this disease is less than 5% (2). Early detecion is urgently needed. Proteomics studies and marker discovery can help for early detection. Currently, the markers for pancreatic cancer are Carbohydrate Antigen 19-9 (CA19-9) and carcinoembrionic antigen (CEA). However, false positives are a problem in the diagnosis (3). Thus, additional diagnostic and prognostic markers are needed. Identification of novel biomarkers may lead to the development of efficacious strategies for early detection and improved therapy.

Protein microarray methodologies have the potential of revolutionizing the analysis of human cancer and have been explored for high-throughput screening of changes in protein expression include both antigen and antibody array formats (4-10). By simultaneously measuring the expression of thousands of proteins in clinical specimens, large numbers of data points can be collected to form a molecular fingerprint of a disease process. However, the traditional natural protein microarray methods sometimes can not provide sensitive immunoreactivity because of the complicated structure of proteins. As shown in Figure 2, the folded proteins may not have the immunoreactivity with the antibody because of the folded antigen. CNBr is a chemical used for protein digestion. It cleaves at Methionine and converts the Met into another amino acid called Homoserine lactone (HSL). Since methionine residues are not as abundant as other residues that are cleaved by popular enzymes, the proteins can be digested to larger pieces. By using digested proteins with larger resulting peptides on the microarray, we can take advantage of potentially immunogenic and biologically significant alterations to the proteins that may still be intact and detectable and increase the chance of the immunoreactivity between antigens and antibodies. The application of this modified protein microarray technology in the study of cancer will help the discovery, characterization, and clinical application of cancer biomarkers.

In this study, we applied a 2-D liquid separation method combined with modified microarray to study the classification of humoral response by exposing them to sera from normal individuals, chronic pancreatic patients and pancreatic cancer patients. This method involves separating intact proteins from cell lysates using chromatofocusing in the first dimension and separating proteins in a second dimension by nonporous silica

RP-HPLC. The result is a 2-D liquid phase fractionation of the proteins from the cell lysate. The fractionated proteins were digested by different digestion methods. The intact proteins and digested proteins can then be spotted on a nitrocellulose slide in parallel. This modified microarray method can be compared to the natural microarray method and used to study humoral response by exposing the arrays to sera from cancer patients, chronic pancreatic patients and normal individuals. Interesting spots which have different immunoreactivity can be identified using LC/MS/MS.

6.2 Experimental Section

6.2.1 Cell Culture and Sample Preparation

6.2.1.1 Chemicals

Acetonitrile, urea, thiourea, DTT, magnesium chloride, glycerol, bis-tris, trifluoroacetic acid, BSA and sodium dodecyl sulfate were obtained from Sigma (St. Louis, MO). Water was purified using a Milli-Q water filtration system (Millipore, Inc., Bedford, MA), and all solvents were HPLC grade unless otherwise specified. Reagents used were in the most pure form commercially available. Polybuffer 74 was purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Anti-human IgG and Alexaflor647 solution were obtained from Invitrogen (Carlsbad, CA). 1X PBS and ultrapure DNase/RNase free distilled water were obtained from Invitrogen Corp. (Carlsbad, CA).

6.2.1.2 Sample Preparation. (a) Cell Culture

The cells used in this work were from the pancreatic adenocarcinoma cell line, Panc-1. The cells were cultured in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum, 100 units/ml penicillin and 100 units/ml streptomycin

(Invitrogen, Carlsbad, CA). When the cells reached ~90% confluence, the cells were harvested with a cell scraper.

6.2.1.2 (b) Cell Lysis

Cell pellets were reconstituted in lysis buffer consisting of 7.5 M urea, 2.5 M thiourea, 4% *n*-octyl-β-D-glucopyranoside (*n*-OG), 10 mM tris(2-carboxyethyl) phosphine (TCEP), 12.5% v/v glycerol, and 1% v/v protease inhibitor cocktail (Sigma, St. Louis, MO). The cell pellets were lysed at room temperature for 1 h, followed by centrifugation at 35 000 rpm at 4 °C for 1 h. The supernatant were buffer exchanged into start buffer (6 M urea, 25 mM Bis-Tris, and 0.2% OG) using a PD-10 G-25 column (Amersham Biosciences, Piscataway, NJ) and stored at -80^{o}C until further use.

6.2.2 Separation

6.2.2.1 Chromatofocusing

CF separation was performed on an HPCF-1D column (250 × 2.1 mm) (Beckman Coulter, Fullerton, CA) using the ProteomeLab™ PF2D protein fractionation system (Beckman Coulter), as described previously (11). Two buffers were used to generate the pH gradient on the column. The SB solution was composed of 6M urea, 25mM Bis-Tris (pH 7.4). The EB solution was composed of 6M urea and 10% polybuffer74 (pH 4.0). Both buffer solutions were brought to pH by addition of a saturated solution of iminodiacetic acid. The CF column was pre-equilibrated with SB. After equilibration, 4.5 mg of proteins were loaded onto the CF column and the column was washed with 100% SB to remove material that did not bind to the column at pH 7.4. Elution was achieved by applying a pH 4.0 elution buffer at a flow rate of 0.2 mL/min. The pH gradient was monitored on-line by a flow-through pH probe (Beckman Coulter). The UV

absorbance of the eluent was monitored on-line at 280nm. The flow rate was 0.2ml/min, with 16 fractions in total being collected in 0.2 pH units in the range of pH 7.0 - 4.0. Each fraction was stored at -80°C until use.

6.2.2.2 NPS-RP-HPLC with sample collection

When the first-dimension separation was completed, the pI fractions collected from the first dimension were separated by nonporous silica reverse phase HPLC (NPS-RP-HPLC) using a ODSIII (4.6 × 33 mm) NPS column (Eprogen, Inc.) and detected by absorbance at 214 nm using a Beckman model 166 UV absorption detector. The RP separation was performed at 0.5 mL/min and monitored at 214 nm using a Beckman 166 Model UV detector (Beckman-Coulter). Proteins eluting from the column were collected by an automated fraction collector (Model SC 100, Beckman), controlled by an in-house designed DOS-based software program. To enhance the speed, resolution, and reproducibility of the separation, the RP column was heated to 65ºC by a column heater (Jones Chromatography, Model 7971, Resolution Systems, Holland, MI). Both mobile phase A: MilliQ® water (Millipore, Billerica, MA), and solvent B: acetonitrile (ACN) (Sigma) contains 0.1% v/v trifluoroacetic acid (TFA). The gradient was run from 5% to 15% in 1 min, 15% B to 25% in 2 min, 25% to 31% in 2 min, 31% to 41% in 10 min, 41% to 47% in 6 min, 47% to 67% in 4 min, then up to 100% B in 3 min where it was held for 1 min, and then reduced to 5% in 1 min. After the gradient, the column was washed by two fast gradients from 5% B to 100% B in 5 min, 100% B back to 5% B in 1 min. Fractions from the HPLC eluent were collected using a semi-automated in-house program using a Model SC-100 fraction collector. Collected peak fractions were stored at -80ºC for further use.

6.2.3 Protein Digest by CNBr

Collected fractions were then dried down and resuspended in 5µL deionized water, 15µL TFA and 5µL 5M CNBr in ACN. Wrapped the tubes in aluminum foil and left them overnight at 4 ºC.

6.2.4 Microarray Printing

Fractionated proteins were transferred to 96-well printing plates (Bio-Rad) and were lyophilized to dryness. The fractions were then resuspended in printing buffer (62.5 mM Tris-HCl (pH6.8), 1% w/v sodium dodecyl sulfate (SDS), 5% w/v dithiothreitol (DTT) and 1% glycerol in 1X PBS) and were left to shake overnight at 4^{o}C. Slides were printed by transferring each fraction from the plate onto nitrocellulose slides using a non-contact piezoelectric printer (Nanoplotter 2, GeSiM). Each spot resulted from deposition of 5 spotting events of 500 pL each, such that a total volume of 2.5 nL of each fraction was spotted. Each spot was found to be ~450 µm in diameter, with the distance between spots maintained at 600 µm. Printed slides were left on the printer deck overnight to dry and were then stored desiccated at 4^{o}C until further use.

6.2.5 Hybridization of slides

The printed arrays were rehydrated in 1X PBS with 0.1% Tween-20 (PBS-T), and were then blocked overnight in a solution of 1% BSA in PBS-T. Each serum sample was diluted 1:400 in probe buffer (5 mM magnesium chloride, 0.5 mM DTT, 0.05% Triton X-100, 5% glycerol and 1% BSA in 1X PBS) to make a total solution of 4 mL and kept on ice. Each diluted serum sample was used to hybridize a slide for 2 hrs. Hybridization was done at 4^{o}C in heat-sealable pouches with agitation, using a mini-rotator. The slides were then washed five times with probe buffer (5 min each), and were then hybridized with 4

mL anti-human IgG conjugated with Alexaflor647 (Invitrogen, Carlsbad, CA) (at 1 μg/mL), for 1 hr at 4^{o}C. After secondary incubation all slides were washed in probe buffer five times, for 5 min each, and were then dried by centrifugation for 10 min. All processed slides were immediately scanned using an Axon 4000B microarray scanner (Axon Instruments Inc., Foster City, CA) and GenePix Pro 6.0 software (Molecular Devices, Sunnyvale, CA) was used for data acquisition and analysis.

6.2.6 Protein Identification

Proteins were trypsin digested as described in other chapters. The tryptic-digested samples were separated by a capillary RP column (C18, 0.3 × 150 mm) (Michrom Biosciences, Auburn, CA) on a Paradigm MG4 micropump (Michrom Biosciences) with a flow rate of 300 nL/min. The gradient, started at 5% ACN, was ramped to 60% ACN in 25 min and finally ramped to 95% in another 5 min. Both solvents A (water) and B (ACN) contained 0.3% formic acid. The resolved peptides were analyzed on A Finnigan LTQ mass spectrometer (Thermo Electron Corp., San Jose, CA) with a nanoESI ion source (Thermo). The capillary temperature was set at 175°C, spray voltage was 2.8 kV, and capillary voltage was 30 V. The normalized collision energy was set at 35% for MS/MS. MS/MS spectra were searched using the SEQUEST algorithm incorporated in Bioworks software (Thermo) against the Swiss-Prot human protein database. The search was performed using the following parameters: one miscleavage is allowed during the database search; Protein identification was considered positive for a peptide with *X*corr of greater than or equal to 3.5 for triply, 2.5 for doubly, and 1.9 for singly charged ions.

6.3 Results and Discussion

6.3.1 Proteome Studies

Figure 1 provides an overview of the experimental procedure. 2-D liquid-based mapping methodology was developed and applied to profile protein expression of Panc-1 cell line. Cells were lysed in a urea buffer system and soluble proteins were separated by their isoelectric point (pI) using chromatofocusing in the first dimension. Equal amounts of protein were loaded for each run and a total of 15 pI fractions corresponding to a pH range of 4.0-7.0 were collected. The second dimension employed nonporous silica reverse-phase high performance liquid chromatography (NPS-RP-HPLC) to separate proteins according to their hydrophobicity. Eluent from the NPS-RP- HPLC step was collected. The collected fractions were separated to three parts, one part as intact protein; other two parts are digested using CNBr or Glu-C. The fractionated proteins or peptides were then arrayed on nitrocellulose slides as parallel unique spots. Hybridization of slides processed with sera from three diagnostic groups, normal sera, chronic pancreatitis and pancreaitc cancer sera. All processed slides were immediately scanned and data were analyzed to find interesting proteins which may serve as potential markers. Mass spectrometry (LC-MS/MS) was used for protein identification.

6.3.2 Reproducibility studies

The reproducibility of the analysis technique was studied. Figure 3A represents the first dimension chromatofocusing profile for four separate runs to explore experimental reproducibility (data only shows three). The second dimension hydrophobicity profiles for fractions 5.2-5.0 and 6.6-6.4 are shown in Fig. 3B. These results indicate high reproducibility between the chromatograms of the band patterns and retention times in both the first dimension chromatofocusing separation and the second dimension reverse phase liquid separation.

6.3.3 Microarray Image analysis

Three sets of experiments were performed to probe the different digestion methods for the presence of tumor-reactive antibodies in the sera of prostate cancer patients, chronic pancreatitis and healthy patients. After 2-D liquid separation, the collected fractions were separated to three parts, one part as intact protein; other two parts are digested to peptides using CNBr or Glu-C. The fractionated proteins or peptides are then arrayed on nitrocellulose slides as parallel unique spots. GenePix Pro 6.0 software was used for data acquisition and analysis. Figure 4 shows a representative scanned image for a slide containing proteins fractionated in the pH range of 5.2-5.0 and 6.6-6.4 which is probed with serum from healthy individuals. The green spots on the slide are the quality control for standards. The immunoreactivity between whole protein and digested proteins were compared in parallel. In the image, we can see that more spots light up for CNBr digest compared to the whole proteins which is consistent with our assumption that some of proteins do not shown the humoral response because of their complicated structures and the digested proteins can help to overcome this problem. All GluC digest spots light up which may due to the remaining GluC in solution which have nonspecific binding to the enzyme.

We compared the levels of antibody binding for each of the fractions among the healthy sample, chronic pancreatic sample and cancer samples in order to identify those fractions most likely to contain cancer antigens. The yellow squares in the figure represent the spot for which immunoreactivity was not observed with this specific fraction when probed with serum from healthy subjects where it was seen for pancreatic cancer serum and chronic pancreatitis serum. This indicated the presence of auto-

antibodies in serum from pancreatic cancer patients and chronic pancreatitis patients against specific cancer associated proteins. These fractions are also marked in Figure 3B as arrows and were identified using mass spectrometry (LC-MS/MS). The following criteria are used for the spots which are picked for these fractions. 1. the signal intensities equal or above 2 times the background 2. equal or above 50% of one group higher than the 2nd highest intensity from the other group are considered. Figure 5 is a representative of the spot in which immunoreactivity was observed in pancreatic cancer patients and chronic pancreatitis patients while not for normal individuals. Scatter and box plots of signal intensities are shown for spot 762. The spot 762 is the fraction from pH 5.2-5.0 and well numbered c6. The intensities of antibody binding are shown left as 1 for healthy patient, 2 in the middle for chronic pancreatitis and right as 3 for the pancreatic cancer. In the plot, each point is an individual sample. The dashed line in the left plot indicates the level of the second-highest healthy sample, which is defined as the threshold. The number 7 in figure indicates there are 7 cancer samples that are above this threshold. For chronic pancreatitis samples, there are 6 samples above this threshold. The box in the right plot indicates the upper and lower quartiles, with the line in the box indicating the median value. Our analysis indicates that there is a higher level of immunoreactivity in the cancer samples and chronic pancreatitis samples as compared to the normal samples for certain fractions and there is not a major difference between the cancer samples and chronic pancreatitis samples.

6.4 Conclusion

In this work, we applied a 2-D liquid separation method combined with a modified microarray approach to study the humoral response by exposing the arrays to

sera from normal individuals, chronic pancreatitis patients and pancreatic cancer patients. The proteins were separated based on their pI and hydrophobicity. CNBr and GluC digested proteins were compared with whole proteins from Panc-1 samples as baits for auto-antibodies in serum. Cancer, Pancreatitis and Normal sera were used for classification of response. Using this method, hundreds of isolated proteins in the liquid phase can be collected and digested for spotting on the microchip array. This method allows for comprehensive analysis of the cancer proteome using small amounts of analyte obtained by fractionation. The modified microarray method using CNBr digested proteins provided more sensitive results for microarray studies compared to the natural whole protein studies. Through the analysis, we found that for some interesting fractions, there is a higher level of immunoreactivity in the cancer samples and chronic pancreatitis samples compared to the normal samples. However, there are not major difference between cancer samples and chronic pancreatic samples. Further studies on fractions of interest may be needed to identify potential serum biomarkers.

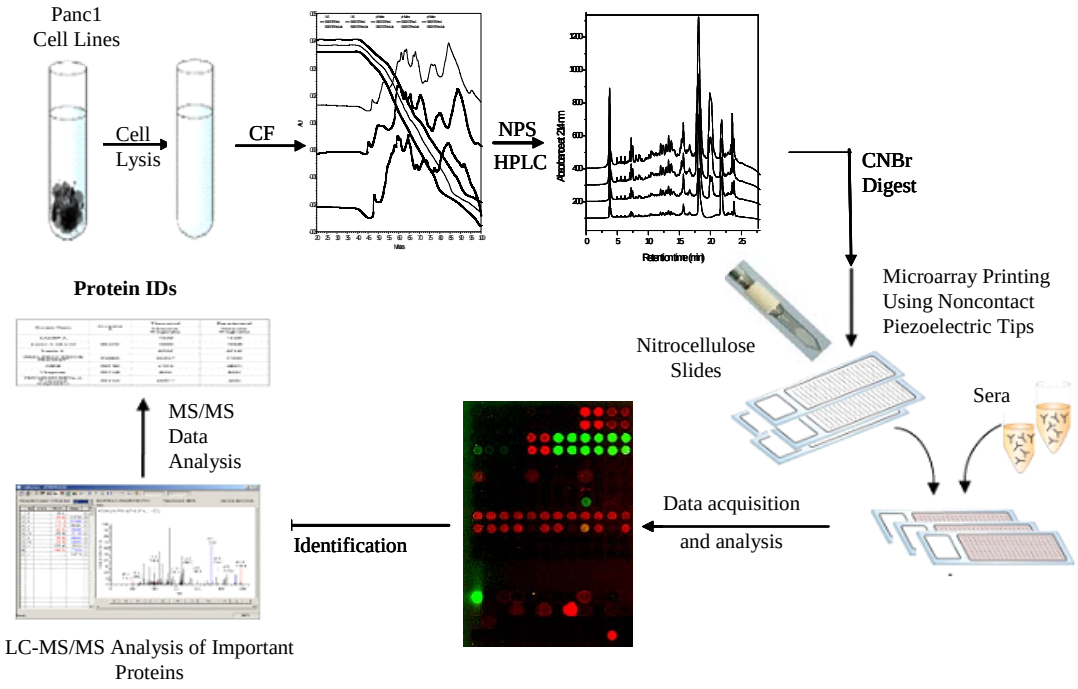


Figure 6.1 Experimental flow chart for modified protein microarry technique

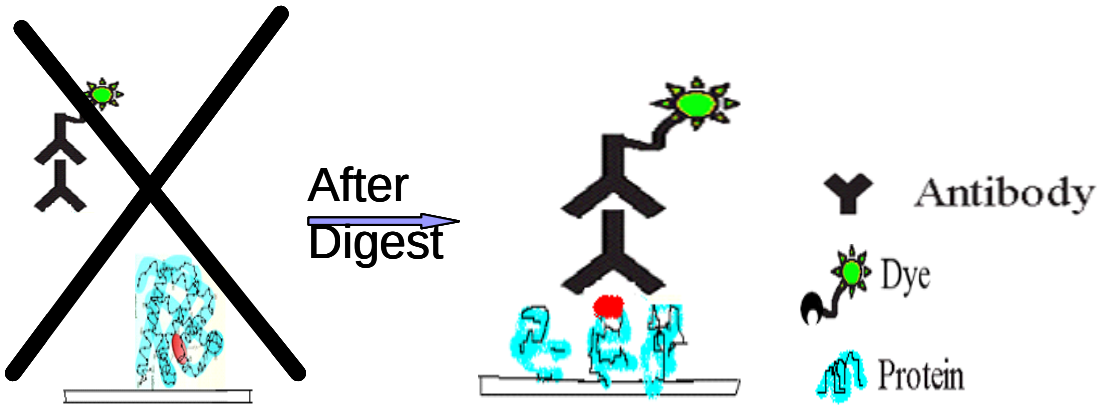
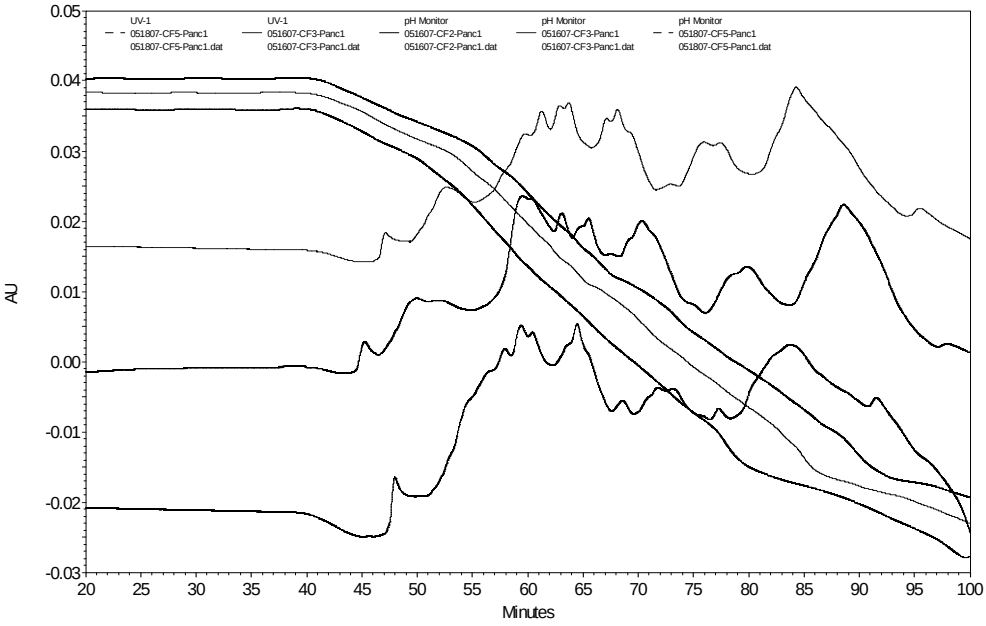


Figure 6.2 Protein arrays for antibody/antigen detection (Humoral Response)

A



B

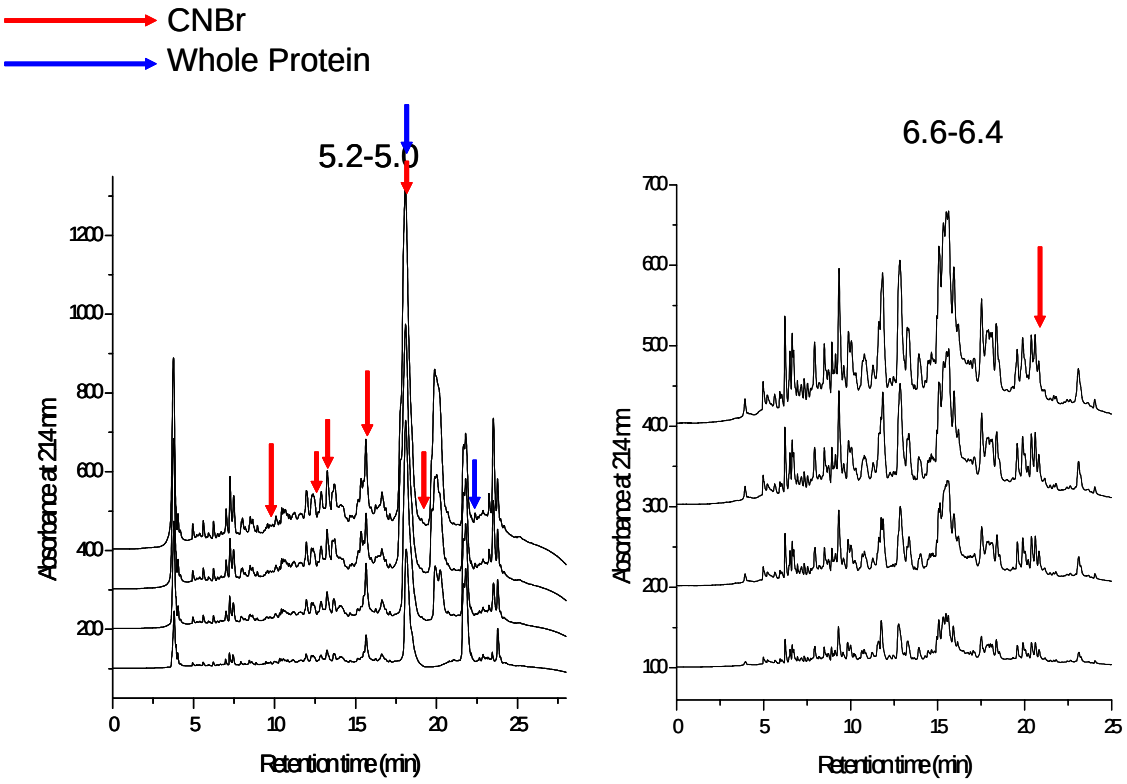


Figure 6.3 Chromatofocusing and HPLC separation.

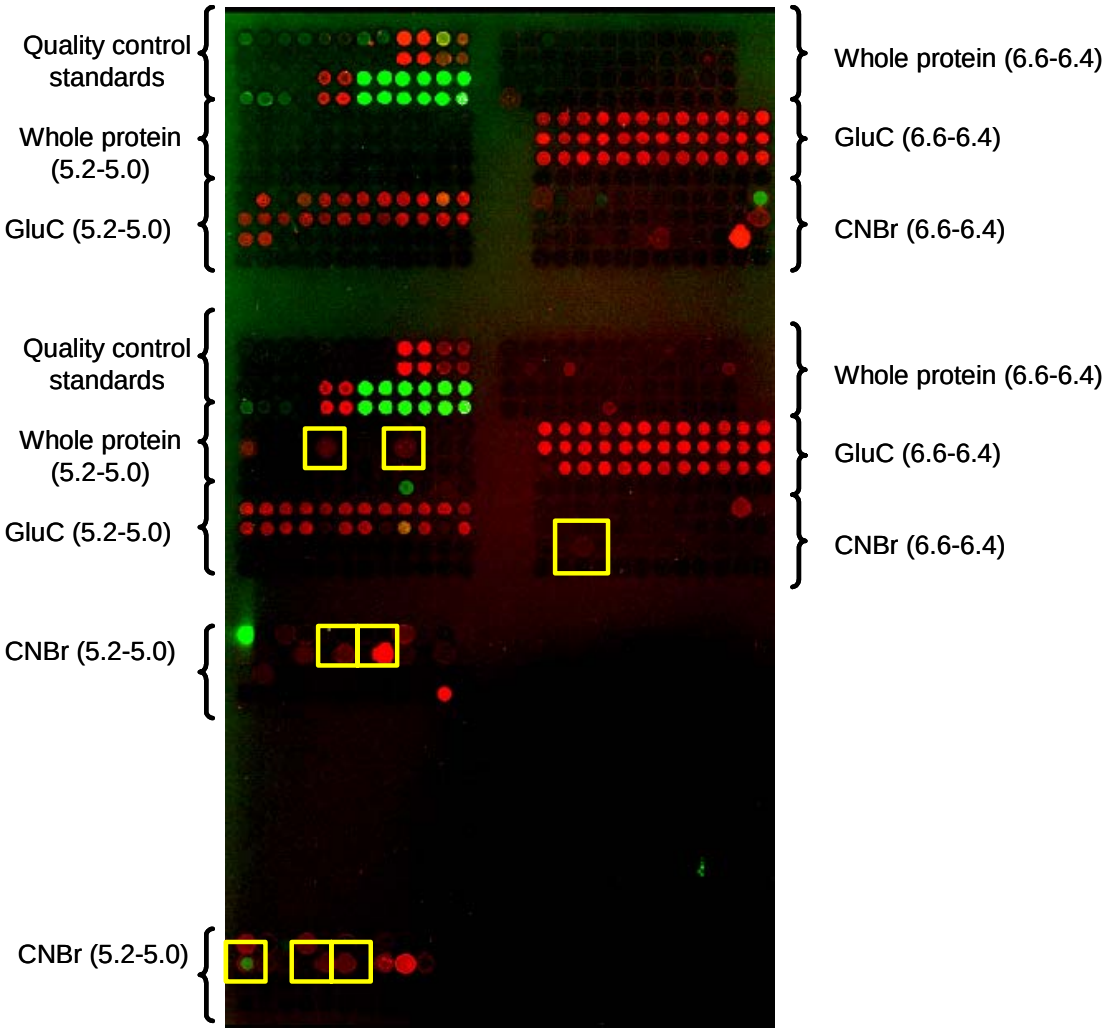


Figure 6.4 Microarray analysis.

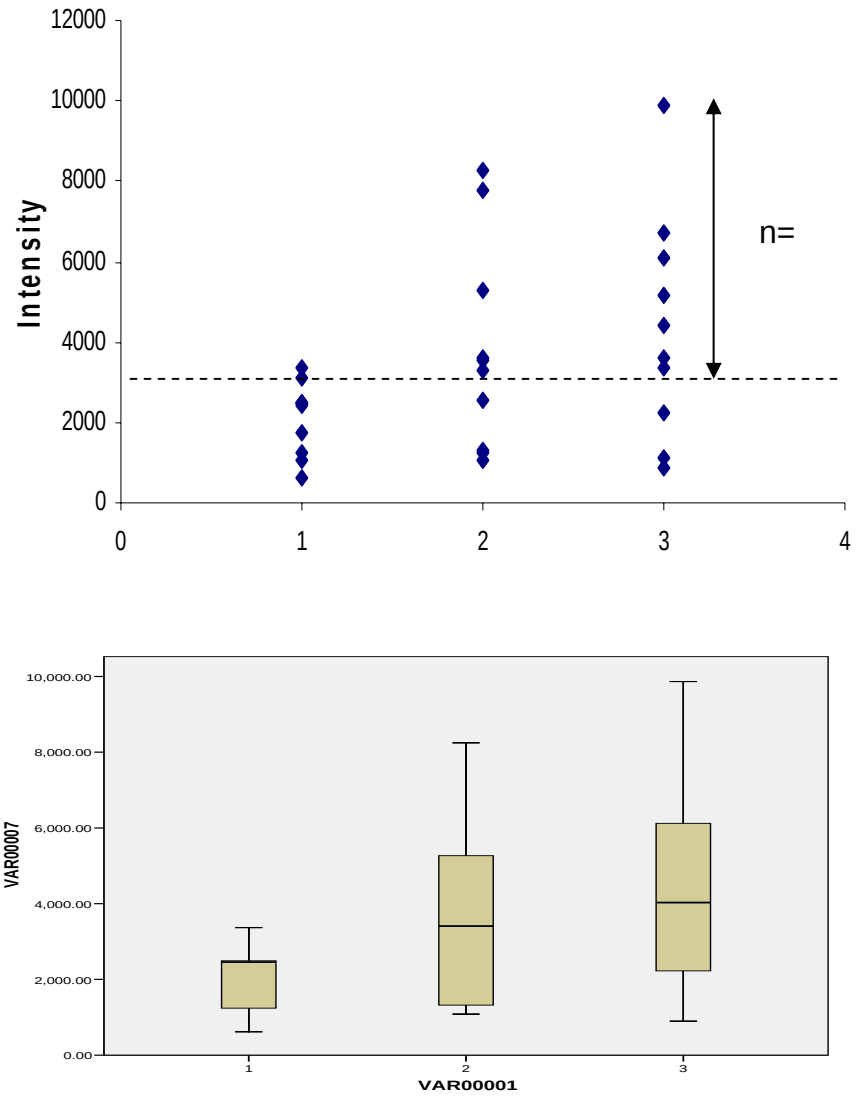


Figure 6.5 Scatter and box plots of signal intensities. 1 healthy patient, 2 chronic pancreatitis, 3 pancreatic cancer. In the plot, each point is an individual sample. The dashed line indicates the level of the second-highest healthy sample, and the number 7 there indicates the number of cancer samples above that threshold. The box indicates the upper and lower quartiles, with the line in the box indicating the median value.

6.5 References:

1. Jemal, A.,Siegel, R.,Ward, E.,Murray, T.,Xu, J. Q.,Smigal, C., and Thun, M. J. (2006) Cancer statistics, 2006 *Ca-a Cancer Journal for Clinicians*, *56*, 106-130.
2. Greenlee, R. T.,Murray, T.,Bolden, S., and Wingo, P. A. (2000) Cancer statistics, 2000 *Ca-a Cancer Journal for Clinicians*, *50*, 7-33.
3. Rosty, C., and Goggins, M. (2002) Early detection of pancreatic carcinoma *Hematology-Oncology Clinics of North America*, *16*, 37-+.
4. Angenendt, P.,Glokler, J.,Murphy, D.,Lehrach, H., and Cahill, D. J. (2002) Toward optimized antibody microarrays: a comparison of current microarray support materials *Analytical Biochemistry*, *309*, 253-260.
5. Sugita, M.,Geraci, M.,Gao, B. F.,Powell, R. L.,Hirsch, F. R.,Johnson, G.,Lapadat, R.,Gabrielson, E.,Bremnes, R.,Bunn, P. A., and Franklin, W. A. (2002) Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma *Cancer Research*, *62*, 3971-3979.
6. Sreekumar, A.,Nyati, M. K.,Varambally, S.,Barrette, T. R.,Ghosh, D.,Lawrence, T. S., and Chinnaiyan, A. M. (2001) Profiling of cancer cells using protein microarrays: Discovery of novel radiation-regulated proteins *Cancer Research*, *61*, 7585-7593.
7. Rimm, D. L.,Camp, R. L.,Charette, L. A.,Costa, J.,Olsen, D. A., and Reiss, M. (2001) Tissue microarray: A new technology for amplification of tissue resources *Cancer Journal*, *7*, 24-31.
8. Li, X.,Mohan, S.,Gu, W.,Miyakoshi, N., and Baylink, D. J. (2000) Differential protein profile in the ear-punched tissue of regeneration and non-regeneration strains of mice: a novel approach to explore the candidate genes for soft-tissue regeneration *Biochimica Et Biophysica Acta-General Subjects*, *1524*, 102-109.
9. Lueking, A.,Horn, M.,Eickhoff, H.,Bussow, K.,Lehrach, H., and Walter, G. (1999) Protein microarrays for gene expression and antibody screening *Analytical Biochemistry*, *270*, 103-111.
10. Martin, B. D.,Gaber, B. P.,Patterson, C. H., and Turner, D. C. (1998) Direct protein microarray fabrication using a hydrogel "stamper" *Langmuir*, *14*, 3971-3975.
11. Wang, Y. F.,Wu, R.,Cho, K. R.,Shedden, K. A.,Barder, T. J., and Lubman, D. M. (2006) Classification of cancer cell lines using an automated two-dimensional liquid mapping method with hierarchical clustering techniques *Molecular & Cellular Proteomics*, *5*, 43-52.

Chapter 7

Conclusion

A two-dimensional liquid mapping method was used to map the protein expression of eight ovarian serous carcinoma cell lines and three immortalized ovarian surface epithelial cell lines. Maps were produced using pI as the separation parameter in the first dimension and hydrophobicity based upon reversed-phase HPLC separation in the second dimension. The method can be reproducibly used to produce protein expression maps over a pH range from 4.0 to 8.5. A dynamic programming method was used to correct for minor shifts in peaks during the HPLC gradient between sample runs. The resulting corrected maps can then be compared using hierarchical clustering to produce dendrograms indicating the relationship between different cell lines. It was found that several of the ovarian surface epithelial cell lines clustered together, whereas specific groups of serous carcinoma cell lines clustered with each other. Although there is limited information on the current biology of these cell lines, it was shown that the protein expression of certain cell lines is closely related to each other. Other cell lines, including one ovarian clear cell carcinoma cell line, two endometrioid carcinoma cell lines, and three breast epithelial cell lines, were also mapped for comparison to show that their protein profiles cluster differently than the serous samples and to study how they cluster relative to each other. In addition, comparisons can be made between proteins differentially expressed between cell lines that may serve as markers of ovarian serous

carcinomas. Some proteins which are differentially expressed between different groups were also identified with MALDI-TOF-MS to obtain finger print information combined with LCT-TOF-MS which provides information about intact molecular weight. The automation of the method allows reproducible comparison of many samples, and the use of differential analysis limits the number of proteins that might require further analysis by mass spectrometry techniques

An alternative 2-D liquid phase mass mapping strategy was also applied to profile protein expression of six ovarian serous carcinoma cell lines. Fractions collected from chromatofocusing every 0.15 pH unit were further separated by NPS-NP-HPLC and detected using on-line ESI-TOF. Fractions were collected and differentially expressed proteins were applied to MALDI-TOF-MS to identify the proteins. This method allows us to separate and map hundreds of proteins in the liquid phase which can be displayed in a 2-D image. For each protein, we can obtain information on the exact intact molecular weight, pI and hydrophobicity. Potential markers for specific groups can be identified. Classification can be also applied to different cell lines based on the mass mapping profile. Mass mapping methods can be used to classify ovarian cancer cell lines and can be used as a fingerprint of the proteome. This method may provide a means of studying proteins in interlysate comparisons over a large number of samples.

The 2-dimensional liquid-based protein mapping method was used to characterize global protein expression patterns in 19 ovarian serous carcinoma tumor samples to facilitate molecular classification of tumor stage. Protein expression profiles were produced, using pI-based separation in the first dimension and hydrophobicity-based separation in the second dimension, over a pH range of 4.0-7.0. A hierarchical clustering

method was applied to protein maps to indicate the tumor interrelationships. The 19 tumor samples could be classified into two different groups, one group associated with low stage (stage 1) tumors and the other group associated with high stage (stages 3/4) tumors. Proteins that were differentially expressed in different groups were selected for identification by MALDI-TOF-MS or QIT-TOF-MS/MS or LTQ-ESI-MS/MS. Fifteen of the selected proteins were over-expressed in the low stage tumors; 49 of the proteins were over-expressed in the high stage tumors. These proteins are known to play an important role in cellular functions such as glycolysis, protein biosynthesis, and cytoskeleton rearrangement and may serve as markers associated with different stages of ovarian serous carcinomas.

In this work, a lectin affinity column was used as an approach to the proteomic analysis of membrane glycoproteins. We compared two different lysis methods and found out that lysis method 1 provides improved results compared to the commercial membrane extract kit. Both cell line samples we did have their own specific glycoprotein expression patterns for both ConA and WGA lectin affinity. We applied the spectral count label-free method for the quantification between CA1a and AT1 and identified differentially expressed proteins between them which may serve as potential markers. We found that N-linked glycoprotein gamma-glutamyl hydrolase only is expressed in the CA1a cell line consistent with the previous references. Further glycoprotein studies on cancer tissue samples could facilitate biomarker discovery. The use of lectin columns in conjunction with mass spectrometry offers a useful approach for the isolation and identification of novel glycoproteins. This work expands our tissue proteome capabilities from the analysis of soluble proteins in previous studies to the examination of membrane proteins.

In other work, a two-dimensional liquid separation technique combined with modified microarray method was applied to study the classification of the humoral response response by exposing the arrays to sera from normal individuals, chronic pancreatitis patients and pancreatic cancer patients. After two dimensional separation according to the pI and hydrophobicity, the protein fractions were collected and different digestion methods were applied. The digested microarray fractions were applied to study the humoral response. The natural microarray method and modified microarray method was compared. We found that CNBr digest provided increased sensitivity compared with the whole proteins. The humoral response for cancer serum and pancreatic serum are different from the normal serum. The potential markers which have different immunoreactivity can be identified using the LC/MS/MS technique.

In summary, the two-dimensional liquid differential mapping method combined with mass spectrometry has been provided a valuable means of analyzing complex biological systems by classifying different groups of human samples and identifying potential markers. This method has several advantages such as the ease of sample handling, automation, accurate molecular weight and better reproducibility for interlysate studies of the protein content of the cells. With the improvement of sensitivity of mass spectrometric instruments and the completion of the human genome sequence, this method has the potential of providing even lower detection limits, improved specificity and higher throughput for analysis of complex biosystems.