COMBINING LASER CAPTURE MICRODISSECTION AND MALDI MASS SPECTROMETRY FOR TISSUE PROTEIN PROFILING: METHODOLOGY DEVELOPMENT AND CLINICAL APPLICATIONS

By

Baogang Jonathan Xu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

May, 2005

Nashville, Tennessee

Approved,

Richard Caprioli

Robert Coffey

Andrew Link

Michael Stone

Dedicated

to my dear parents, Deshi Xu and Mingrong Jiao,

and

to my lovely wife, Angel Qi An

ACKNOWLEDGEMENTS

I am deeply grateful to numerous people for their support throughout my graduate study. First, I would like to express my gratitude to my thesis advisor, Dr. Richard Caprioli, who has provided me with superb guidance, inspiration, and support. The training he has provided me, both as a critical thinker and innovative scientist, will have life-long benefits. Thanks also to the other members of my committee, Drs. Robert Coffey, Andrew Link, and Michael Stone, for their encouragement and advice concerning my graduate research and scientific career.

I am also thankful for the excellent collaborative environment at Vanderbilt University. I have gained so much from my collaborators. I appreciate Drs. Shyr Yu, Agnes Fogo, Melinda Sanders, Bob Whitehead, David Carbone, Hal Moses and their lab members for their great collaborative efforts. Many thanks to the past and present members of the mass spectrometry research center. My experience has been truly enjoyable at the MSRC. I wish to thank my colleagues for their invaluable technical advice, discussions and overall support. Funding for my research was provided by NHI/NIGMS GM 58008-05 and 2U01 CA842390-06.

I would also like to thank all of the teachers who taught me throughout my education, both in China and in the U.S. I am thankful for Beene's family for their support and love over the years. I will be forever grateful to my parents for their love and support. I credit them with a major part of my every success. I am deeply thankful for my dear wife, Angel An, whose patience, love, and caring give me courage and confidence to face all of life's challenges. I am so blessed to have her in my life. Finally, I would like to thank my lord and savior, Jesus Christ, for all of his provision. May my humble research reflect his infinite glory.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2-D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
2-D DIGE	Two-dimensional difference gel electrophoresis
ADH	Atypical ductal hyperplasia
APC	Adenomatous polyposis coli
CHTN	Cooperative human tissue network
CRC	Colorectal cancer
CRIP	Cysteine-rich intestinal protein
DCIS	Ductal carcinoma *in situ*
ECM	Extracellular matrix
ESI	Electrospray ionization
FAP	Familial adenomatous polyposis
FSGS	Focal segmental glomerulosclerosis
GEN	Glomerular endothelial
GRX	Glutaredoxin
H&E	Haematoxylin and eosin
HNPCC	hereditary non-polyposis colon cancer
HRP	Horseradish peroxidase
ICAT	Isotope-coded affinity tag
IMC	Invasive mammary carcinoma
LCM	Laser capture microdissection
MALDI	Matrix-assisted laser desorption/ionization
MCD	Minimal change disease
Min	Multiple intestinal neoplasia
MudPIT	Multidimensional protein identification technology

NCBI	National Center for Biotechnology Information
PAI-1	Plasminogen activator inhibitor 1
PMSF	Phenylmethanesulfonyl fluoride
RM	Reduction mammoplasty
SAM	Significance Analysis of Microarrays
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SH3BGRL3	SH3 domain-binding glutamic acid-rich-like protein 3
Spink3	Serine protease inhibitor kazal type 3
TFA	Trifluoroacetic acid
TOF	Time-of-flight
t-PA	Tissue-type plasminogen activator
u-PA	Urokinase-type plasminogen activator
VEGF	Vascular endothelial growth factor
WFCCM	Weighted flexible compound covariate method
WGA	Weighted Gene Analysis

CHAPTER I

BACKGROUND AND OBJECTIVES

The Era of Proteomics and Biomarkers Discovery

Proteomics encompasses the comprehensive identification and quantification of proteins specific to the physiological state of a cell, tissue or organism. The term "proteome" or "proteomics" was introduced in the mid-1990's, although proteomic strategies have been around for decades [1]. Instead of focusing on one particular protein, proteomic studies often take the approach of investigating hundreds or even thousands of proteins simultaneously in a single experiment by coupling newly developed proteomic strategies with modern mass spectrometers. [2-4].

Proteomic studies are complementary to genomic studies, although the proteome is far more complex than the genome for several reasons. First, one gene may result in multiple protein products due to gene splicing [5]. Second, posttranslational modifications are essential for protein structure and function, which cannot be fully predicted from the genome sequence [6]. Third, the proteome is highly dynamic in relative abundance and subcellular localization. All of these aspects should be considered when investigating the physiological state of a cell or tissue.

As a technology-driven science, proteomics has been accelerated by the advancement of various related research areas. First, the completions of the genomic sequences of human and other species have, in principle, provided the amino acid sequence of all the potentially encoded proteins for those species [7, 8]. The resulting genomic sequence databases have provided a library of amino acid sequence information that is required for proteomic study. Second, the development of modern mass spectrometry, especially matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) mass spectrometry [9, 10], has made it possible for

the ionization of nonvolatile, large analytes such as proteins without significant fragmentation. MALDI and ESI mass spectrometry have emerged as very important instrumentations for protein identification, posttranslational modification characterization, assessing protein expression pattern, solving protein-protein interaction and protein structure [11-13]. Third, the emergence and development of bioinformatics helps to integrate the vast amount of information obtained from genomic and proteomic studies to assemble comprehensive and biologically relevant molecular pathways. The data management, mining and interpretation for proteomic studies are made possible as a direct result of the advancement in bioinformatics [14, 15].

As a discovery-driven technology, the scope of proteomics extends beyond simple cataloguing of proteins within an organ to encompass detailed functional analysis of proteins. Proteomics approaches have been widely utilized to classify and elucidate mechanisms underlying various diseases, such as cancer, cardiovascular disease, and various neurological disorders [16-19]. A detailed understanding of proteins found to be differentially expressed in a disease-specific fashion would be expected to offer great insight into the understanding of specific pathogenic mechanisms, thereby providing diagnostic markers as well as potential drug targets. From this perspective, recently discovered protein markers derived from proteomic studies have been suggested as having the potential for diagnostic and prognostic applications in the areas of ovarian, lung and brain tumors [20-23].

Proteomics Technologies

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is one of the most widely used techniques for proteomic studies. It is a powerful protein separation technique which can resolve up to thousands of different protein species [24]. The protein mixtures are first separated according to their isoelectric point by isoelectric focusing and then separated according to their size by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The

separated proteins are visualized with different staining techniques, such coomassie blue staining. Differentially expressed proteins are identified by comparing spot intensity between samples. The gel spots of interested are excised and digested with a protease. From the digested peptides, the identities of these protein markers are characterized using peptide mapping or peptide sequence analysis using mass spectrometry. This technique, however, has some limitations. Very acidic or basic proteins, very large or small proteins, and membrane proteins are underrepresented in 2D PAGE pattern. The technique is also time consuming and requires a relatively large sample size, usually in the microgram range. The gel-to-gel variance can be significant but the introduction of two-dimensional difference gel electrophoresis (2-D DIGE) technology, which allows for the direct comparison of protein abundance changes across multiple samples simultaneously in one experiment, has minimized the gel-to-gel variation [25].

Liquid chromatography-based peptide separation coupled with tandem mass spectrometry offers another powerful technique for proteomic studies [2, 4]. In this approach, complex protein mixtures are directly digested and resulting peptides are separated over multiple liquid chromatographies based on different peptide physical properties. A commonly used approach is referred to as multidimensional protein identification technology (MudPIT). This approach involves separation by strong cation exchange chromatography followed by subsequent separation of each corresponding fraction with reverse phase chromatography using capillary HPLC [2, 26]. The separated peptides are directly introduced into the mass spectrometer after eluting off from the reversed phase column. The downstream tandem mass spectrometry is operated in a data-dependent acquisition mode. Computer algorithms such as SEQUEST can then be used to determine the "best fit" between the MS/MS sequence information and the protein databases. In this way, more than one thousand proteins are identified within a single MudPIT experiment [27]. One limitation of these analyses is that they generally are qualitative while quantification analysis is critical for comparative proteomics. One possible solution to this is the

various labeling techniques that have been developed for semi-quantitative analysis using LC-MS/MS, such as isotope-coded affinity tag (ICAT) technology [28]. The ICAT approach reduces the sample complexity by purifying only the cysteine-peptides and provides a promising tool for the relative quantification of low abundant proteins. It has been successfully applied to various comparative proteomic studies [29, 30].

Direct Tissue Protein Profiling Using MALDI Mass Spectrometry

Direct tissue protein profiling and imaging using MALDI mass spectrometry has emerged as a valuable discovery tool to visualize specific proteins from intact tissue sections. Upon direct analysis of tissue sections cut from frozen tissue using MALDI MS, hundreds of peptides and proteins can be simultaneously detected and mapped [13, 31-33]. Matrix solution, such as sinapinic acid, is directly deposited onto tissue sections. The soluble proteins are extracted from the tissue and co-crystallize with the matrix. Protonated protein ions are formed by irradiating the matrix with a pulsed laser and are separated by mass to charge in a time-of-flight (TOF) mass analyzer. Protein species in the mass range of 2,000-70,000 Da are usually detected in the mass spectra, although protein signals with molecular weights well above 200 kDa have been observed [34]. Matrix solution can be deposited at different regions on the tissue section depending on the histology to produce distinct protein profiles from the specific regions of interest. Figure 1 shows an example of a frozen tissue section with matrix spots on a MALDI plate. The mass spectra illustrate the different protein profiles observed from a normal mouse breast and mouse breast tumor (Figure 1). Generally, the matrix spot and MALDI laser diameter determine the spatial resolution of the tissue protein profile. The spot size of commercially available instruments is typically focused to 60~150 micrometers in diameter. The conventional matrix spot delivered using a commercially available pipette is generally greater than 500

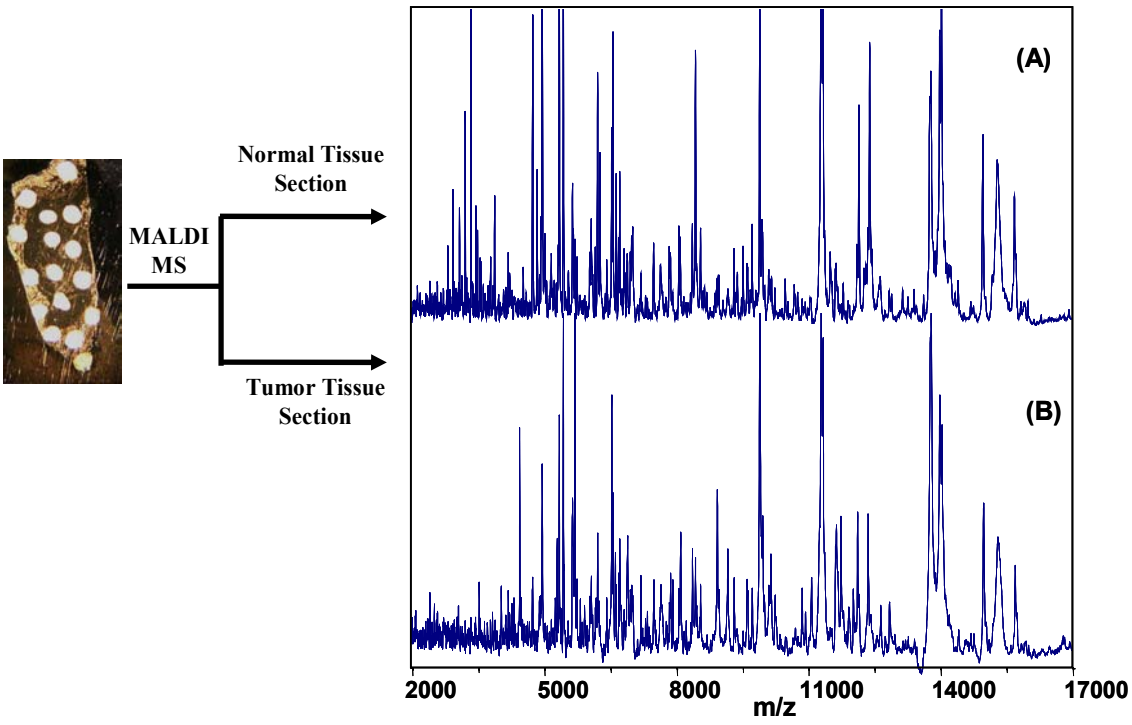


Figure 1: Direct protein profiling from mouse breast tissue by MALDI MS. The picture on the left shows a tissue section with matrix solution spots. The mass spectrum on the top shows the protein profiles from a normal mouse breast tissue (A) and the spectrum at the bottom shows the protein profiles from a mouse breast tumor section (B).

micrometers in diameter. This spatial resolution limits the analysis of minute tissue structure making single cell analysis hard to achieve.

Although various proteomics approaches have been implemented in protein biomarker discovery, the sensitivity and specificity of the biomarker need to be improved. The technologies described above mainly detect the most abundant proteins from the sample. The less abundant or membrane proteins, which are not routinely detected using the current proteomic technologies, could be very important for understanding the mechanisms underlying the disease. Improving the specificity of the protein biomarker is necessary for various clinical applications. Obtaining protein biomarkers which are specific for a disease or different stages of a disease would greatly benefit the diagnosis and prognosis for patient care.

Laser Capture Microdissection

The Necessity of Tissue Microdissection

The heterogeneous nature of tissue presents a great challenge for genomic and proteomic research. When DNA, mRNA, or proteins are analyzed from a bulk of tissue, it is hard to assign the molecular markers of interest to any specific cell type. Identification of disease markers associated with the diseased cells is hindered by the complexity of tissues. The target cells, for example, breast tumor cells, often infiltrate irregularly into tissue and neoplastic cell are surrounded by large amounts of stroma, lymphocytes, blood vessels, etc, as shown in figure 2. Without sampling a homogenous tumor cell population, even significant differences resulting from the comparison of control and diseased cells could be caused by the inherent differences between different types of cells. For example, epithelial cells and stroma cells have different protein expression profiles because of different cellular functions. Furthermore, many minute cellular structures require special techniques to enrich a sufficient amount of sample prior to

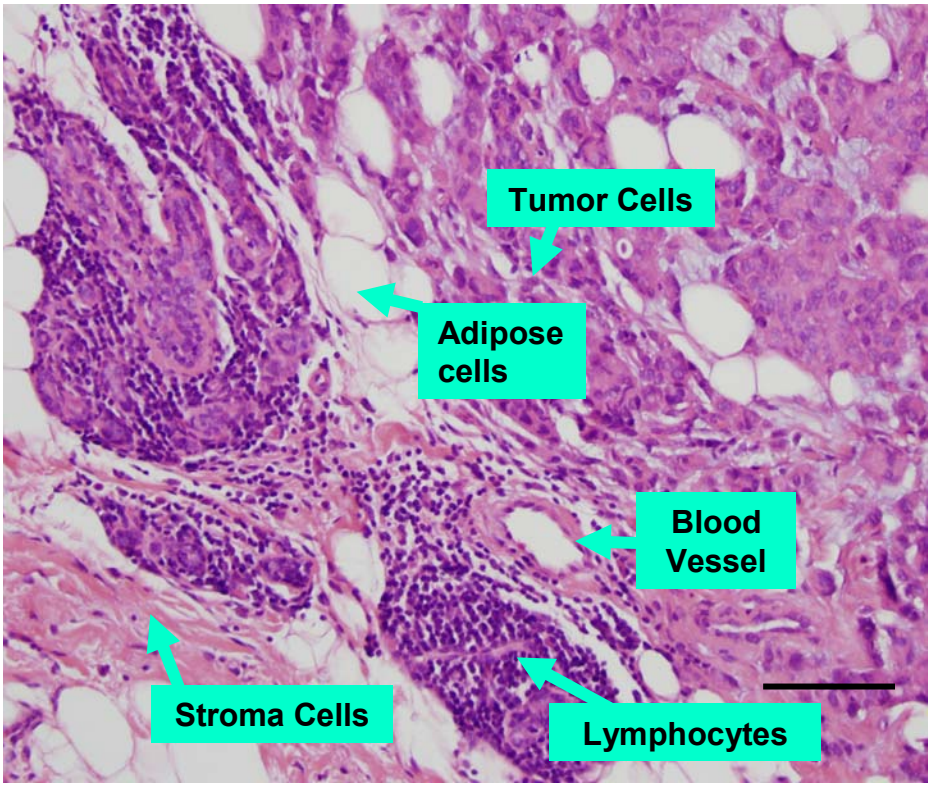


Figure 2: The heterogeneous nature of tissue. The breast tumor cells are surrounded by different types of cells in a tissue section. The scale bar is 100 micrometer.

analysis. Rat kidney glomeruli, for example, are vital structures that filter blood in the kidney. They are less than one hundred micrometers in diameter and are scattered throughout the kidney cortex. A special technique capable of harvesting these cells is essential to assess their molecular changes in different disease states.

Different Laser Microdissection Techniques

The need for obtaining a pure cell population for DNA, RNA and protein analysis in a heterogeneous tissue has driven the development of various cellular microdissection techniques. Although manual microdissection is possible [35], the reproducibility and high throughput are limited. Negative ablation of surrounding tissue of target cells has been used [36, 37], but a significant amount of valuable tissues are destroyed and the process is slow and tedious. More recently, two laser-assisted microdissection techniques have been developed for both paraffin and frozen tissue sections. One approach is called PALM, an acronym for Positioning and Ablation with Laser Microbeams, developed by PALM Microlaser Technology AG (Bernried, Germany) in 2000. In the PALM technique, tissue sections are first dehydrated through a series of alcohol and xylene washes. The procurement of target cells requires two steps: laser ablation and laser pressure catapulting. First, a highly focused beam laser (~ 3 micrometers in diameter) precisely ablates the outline of the selected cells. Thus, the cells within the circles are detached from the rest of the tissue, which yields a clear cut gap between selected and non-selected areas (Figure 3A). Second, using the laser pressure catapulting technique, the isolated cells are ejected from the tissue section into a microtube cap placed above the tissue (Figure 3B). The cap is usually filled with fluid, such as culture medium, to catch the ejected cells. Since this liquid medium has to be held in the cap upside down, the medium is required to have a high surface tension. After the cells are collected, the solution in the cap is centrifuged down into a microtube and mRNA or proteins are extracted out of the cells for downstream analysis. Although the technique is effective, several

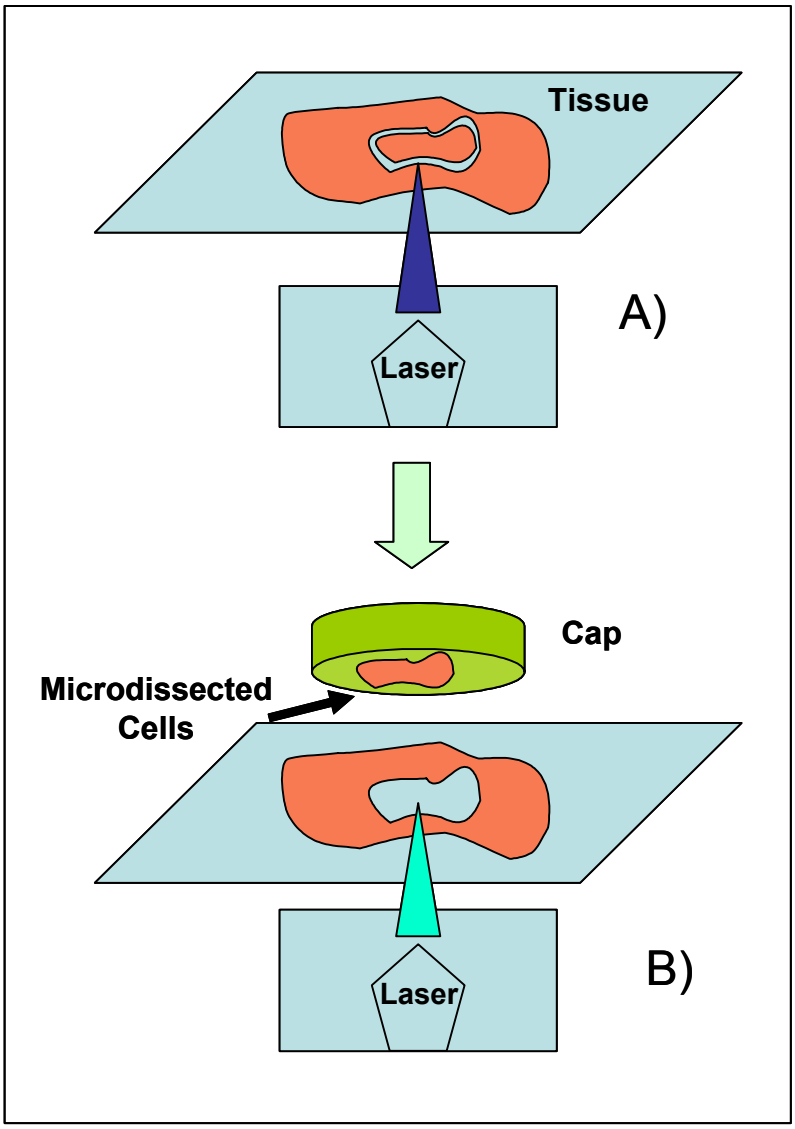


Figure 3: The process of P.A.L.M. laser microdissection process. A) The laser beam precisely cuts an outline for the selected area on the tissue, which yields a clear cut gap between selected and non-selected areas. B) After microdissection, the isolated specimens are ejected from the tissue section and catapulted directly into the cap of a common microfuge tube.

disadvantages limit its wide application. The necessity and special requirements for the extraction medium dilutes the protein contents and hinders some downstream analysis techniques, such as the direct protein profiling using MALDI MS. The complex operational procedure and extensive instrumental maintenance have also contributed to its lack of popularity. Only limited applications have been reported using this technology, with most of these focused on measuring DNA or RNA levels [38, 39].

Laser Capture Microdissection Mechanisms and Applications

The second laser assisted microdissection technique was first described in 1996 by Michael Emmert-Buck *et al*, in which a laser was used to capture the specific cells of interest onto a thermoplastic film [40]. Named laser capture microdissection (LCM), this technique was commercialized by Arcturus Bioscience Inc. (Mountain View, CA) and has become a widely used laser microdissection technique.

The principle of LCM is illustrated below. The tissue section is mounted on a regular histology glass slide and dehydrated with a series of alcohol and xylene washes. The glass slide is placed on a microscope stage with an inverted lens. As shown in figure 4, a transparent cap with thermoplastic film is placed on the top of the tissue section. A low power infrared laser beam is directed at the cells from above. This special film at the bottom of the cap is made of thermoplastic polymer mixed with proprietary metal. The histology of the tissue can be visualized under the microscope, and the specific cells of interest are positioned under the laser using a handle. When the operator activates the laser, the thermoplastic film is transiently melted above the target cells. The long polymer chains from the film infiltrate into the tissue and tightly bind with the selected cells underneath. The adhesion force between the film and cells separates the target cells from the surrounding tissue when the cap is lifted. The selected cells are embedded in

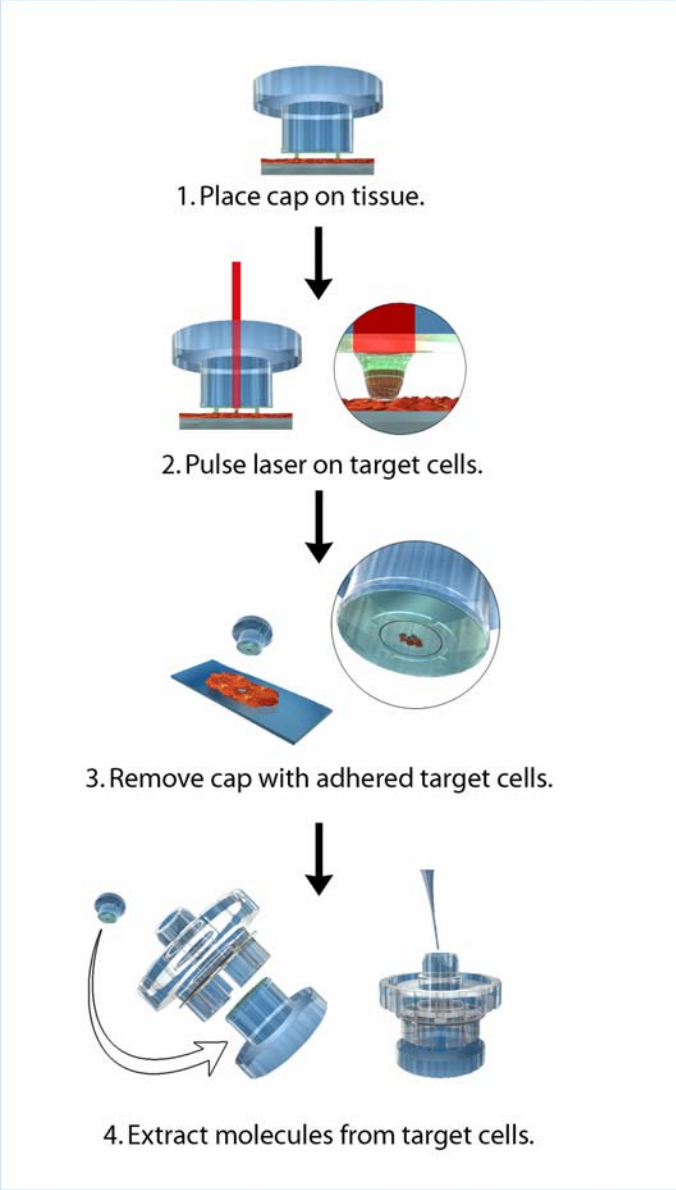


Figure 4: The laser capture microdissection process. (Courtesy of www.arctur.com)

the film which can then be placed into a buffer for extraction of either DNA, RNA or proteins for subsequent analysis.

LCM has been readily utilized for collecting pure cell populations for gene microarray analysis. Real-time PCR technology can amplify the mRNA from a very small amount of sample prior to analysis. The coupling of real-time PCR and LCM have been applied to many areas in the biomedical science, such as cancer research [41-43], neurological disease [44], proteinuric disease [45], HIV disease [46], and studies of hormonal receptors [47]. However, unlike mRNA, proteins cannot be amplified after the sample collection, thus the limited number of cells obtained using LCM technique cannot be easily applied to traditional proteomic analysis. For example, 50,000 to 100,000 cells are required for one 2D-PAGE experiment [48, 49]. Obtaining this large amount of cells manually is very labor-intensive, and cannot be done routinely and limits any high throughput analysis. Therefore, combining LCM with the direct tissue analysis using MALDI MS offers an attractive approach for obtaining protein profiles from a small population of highly selected cells.

Significance of Method Development

Since both LCM and direct tissue protein profiling using MALDI MS offer unique strengths, combining these two techniques would be expected to have a significant impact on proteomic research. After the specific cells of interest are microdissected on the LCM film, instead of immersing the sample in protein extraction buffer, the LCM film with the microdissected cells can be mounted on a MALDI target plate. The protein contents from the captured cells can be analyzed directly using MALDI MS after applying matrix solution. Although this concept was proposed previously by Palmer-Toy *et al* [50], only a few protein signals with low resolution were detected from the cells obtained by LCM. The application of this technology is difficult without dramatically improving the sensitivity. Several protocols of the

LCM methodology need to be optimized and various aspects of the method need to be evaluated. With abundant protein signals detected, the mass spectra obtained will better reflect the protein expression profiles of the microdissected cells. Combining downstream statistical analysis and protein identification, this approach will offer a unique and powerful technology for protein marker discovery.

This seemingly straight forward method of combining LCM and MALDI MS has multiple advantages compared to the current proteomic approaches, such as 2-D gel electrophoresis. First, the protein expression profiles are acquired only from cells of interest. The pure cell population obtained using LCM minimizes the risk of mixing inherent differences existing among different types of cells into the final results; therefore, the protein profiles comparison is much more relevant and meaningful because cells from the same origin are compared. Second, this method is very sensitive because of the high sensitivity of MALDI-TOF MS. Protein profiles can be obtained directly from only a few cells, which dramatically reduces the amount of time and effort required in the microdissection step. Studies which have only little amounts of materials are possible with this technique. Minute tissue structures can be analyzed directly without sample enrichment. Third, the time required for obtaining protein profiles is very short. After the cells are microdissected, the protein profiles can be generated within 20 minutes. Compared to the days 2-D PAGE takes, this high throughput method can be applied to a large cohort of samples to achieve statistically significant results. For example, human tumor biopsies are very heterogeneous due to environmental and genetic variabilities. Analyzing a large number of samples is essential to achieve statistically significant results. Fourth, relatively small proteins and peptides which cannot be easily analyzed by 2-D PAGE are accurately measured using this technique. This offers a great complementary technique for the 2-D PAGE and multidimensional LC-MS/MS approaches.

Research Objectives

The technical protocols involved in combining LCM and MALDI MS for protein profiling of cells must be optimized and evaluated. It is expected that the application of this methodology to various biological systems will aid the discovery of relevant disease-specific protein markers. Furthermore, this approach can be applied to the discovery of new neoplastic tumor markers, new tumor classification, elucidation of pathomechanisms, and prediction of disease progression. Therefore, the following specific objectives are formed for this research:

Objective 1: Develop a reliable and reproducible method for the direct acquisition of protein expression profiles from the cells obtained by LCM using MALDI MS.

Objective 2: Apply the method developed in Objective 1 to various biological systems for the purpose of:

1) Identifying protein markers differentially expressed in normal and neoplastic mouse colon.
2) Obtaining proteomic patterns to begin classifying glomerulosclerosis at the protein level and to advance the understanding of glomerulosclerosis mechanisms.
3) Classifying human breast tumor from normal breast tissue and begin to distinguish *in situ* from invasive cancer at the protein level.

CHAPTER II

METHODOLOGY DEVELOPMENT

The combination of LCM and MALDI MS is an ideal discovery tool to obtain protein profiles from a pure cell population sampled in a heterogeneous tissue [50, 51]. Although holding great promise in the proteomic field for various biological systems, there are several technical challenges to overcome for this technology to become a reliable and sensitive technique. First, a system needs to be developed to accurately deliver matrix solution to the small number of cells captured on the LCM film. LCM has the potential of capturing a single cell, which usually is approximately 10 micrometers in diameter. Regular pipettes cannot deliver the small amount of matrix solution needed to just cover the captured cells. Excess matrix dilutes the proteins and decreases the MALDI MS sensitivity. Second, conventional used histological staining, such as haematoxylin and eosin, may suppress the protein ionization process affecting the mass spectra sensitivity, mass accuracy and reproducibility. Histological staining that is compatible with LCM and MALDI MS needs to be explored. Third, the ethanol and xylene dehydration process required by LCM can potentially wash away some of the highly soluble proteins from the tissue. Therefore, the dehydration effects on protein profiles need to be evaluated. Fourth, the effects of the LCM polymer film on MALDI MS analysis need to be assessed.

The procedures for sample preparation for LCM and MALDI MS analysis must be optimized to achieve highly sensitive and reproducible results required for tissue protein profiling. For this reason, the practical aspects of tissue section preparation, the LCM process, matrix solution deposition, and MALDI mass spectra acquisition are described in this chapter. The effects of different histological staining, alcohol dehydration and the LCM thermoplastic film on MALDI MS analyses are also evaluated and discussed.

Tissue Preparation for Laser Capture Microdissection

Tissue Sectioning and Dehydration

Proper tissue sectioning and dehydration are important prerequisites to achieve optimal results using LCM. We have found previously that surgically procured tissue samples should be snap-frozen in liquid nitrogen and stored in a liquid nitrogen tank or a –80°C freezer prior to use [34]. Five micrometer thick sections are generally cut using a cryostat and mounted on regular histological glass slides without the contamination of embedding media, such as OCT. Sections of greater thickness can also be cut for LCM in an instance where the tissue sample is too small or the structure is too loose to be cut at five micrometer thickness. Formalin fixed tissue samples are not suitable for downstream MALDI MS analysis due to the protein cross linking; therefore, only frozen tissue sections are used in this study.

Complete dehydration of the tissue sections is an important procedure because this ensures an effective transfer of cells to the LCM cap. The presence of water within the tissue section inhibits good capture of the cells. The dehydration should be started immediately after the tissue is mounted on the glass slide. Delaying dehydration will cause the tissue section to adhere to the glass slide irreversibly, and the cells cannot be microdissected. Briefly, the sections were processed as follows: 70% ethanol for 30 second, 95 % ethanol for 1 minute, 100% ethanol for 1 minute (times 2), xylene for 2.5 minutes (times 2), and air dry for 5 minutes.

The Laser Capture Microdissection Process

Prior to laser capture microdissection, the Arcturus Prep-strips (Mountain View, CA) are applied to the dehydrated tissue sections to remove any attached loose particles, such as tissue debris. The glass slide is placed on the LCM microscope stage and one CapSure cap is gently placed on top of the tissue section. The laser from the LCM instrument is focused and the laser

power is optimized. A round shaped ring with a black circle is observed when the laser is fired on the LCM film as an indication for a well focused laser beam. Three different laser spot sizes can be chosen depending on the microdissection resolution requirement, as shown in Figure 5. Cells cannot be captured if the laser is not well focused or the laser power is not high enough. The laser duration time is adjusted according to the spot size required. Table 1 lists the reference range for the laser power and duration time for different laser spot sizes. When the laser is activated the film softens, flows out and adheres to the target cells. The film resolidifies with the adhering cells attached to it. When the cap is lifted, the cells are microdissected from the surrounding tissue.

After the microdissection, the cap is removed from the microscope stage and enclosed in an eppendorf tube to avoid contamination. If MALDI MS analysis cannot be performed on the same day, the LCM caps are stored at -80°C freezer. Prior to MALDI analysis, the LCM caps are warmed to room temperature in a vacuum desiccator to avoid moisture condensation.

Mounting the LCM Film on MALDI Plate

After the LCM cap is removed from the eppendorf tube, the film at the bottom of the cap is gently peeled off using a pair of fine tweezers. A piece of double-sided conductive tape (Digi-Key, Thief River Falls, MN) is mounted on a polished, gold-coated MALDI target plate. The LCM film is mounted flat on the conductive tape with the captured cells facing upward (Figure 6A). An Arcturus prep-strip is applied to ensure no air bubbles exist under the film. Multiple LCM films may be placed on one MALDI plate.

Matrix Solution Deposition

Due to the small number of cells captured on LCM film, conventional pipettes cannot deliver matrix solution droplets small enough to cover only the cells. Excess matrix solution deposited with a conventional pipette dilutes the cellular contents and the resulting large matrix

crystal makes it difficult to locate the cells. The MALDI mass spectral quality is therefore degraded and the sensitivity of this technique suffers. To improve the sensitivity, micro-spotting only appropriate amounts of matrix solution needed to cover the captured cells is a key step.

Though multiple techniques are available for small droplet deposition, a special approach has to be taken for this research because the sinapinic acid matrix solution crystallizes and clogs the line tips very quickly. Utilization of a special manually controlled pulled glass capillary is a good approach for several reasons: 1) the opening tip of the capillary can be cut with a 10 micrometer diameter and used to deliver sub-nanoliter volumes of solution to cover the captured cells. 2) Manually controlling the capillary enables the deposition in a fast fashion, thereby avoiding the matrix clogging problem. 3) The glass capillary is low cost and disposable. Changing a new capillary for each sample is practical and avoids possible cross contamination. A new capillary can be used if it gets clogged. 4) The surface tension of the glass capillary automatically draws in matrix solution. The matrix solution can be deposited onto target cells by touching the capillary tip to the captured cells.

The matrix solution is deposited onto the captured cells under microscope visualization. The matrix solution, typically consisting of 20 mg/ml sinapinic acid in 6/4/0.03 v/v/v acetonitrile/water/TFA, is drawn into the capillary by dipping the tip into the matrix solution for several seconds. The MALDI plate with the LCM film is placed on the microscope stage. The captured cells can be clearly visualized under 4X object lens with proper light from above. The capillary tip loaded with matrix solution is gently touched to the surface of the cells (Figure 6B) and the matrix solution is deposited onto the cells. The matrix solution forms crystals several seconds after deposition on the cells (Figure 6C). The volume of matrix solution deposited ranges from approximately 100 pl to 10 nl depending on the number of microdissected cells.

Figure 5: The expected shape for a well focused laser spot on a LCM film. Three different laser spot sizes were presented. The scale bar is 30 micrometer.

Table 1: The reference ranges of laser power and duration for focusing LCM laser at different spot sizes. These specifications are for the CapSures caps from Arcturus Co. (Courtesy form Arcturus Co.)

LASER SPOT SIZE	LASER POWER	LASER DURATION
Small (7.5 μm)	40-50 mW	550-650 μs
Medium (15 μm)	30-40 mW	1.5-2.0 ms
Large (30 μm)	25-35 mW	5.0-6.0 ms

Analyzing the Laser Capture Microdissected Cells by MALDI MS

Sensitivity

To evaluate the sensitivity of this technology, various numbers of cells were microdissected by LCM and analyzed using MALDI MS. For comparison, 5, 20, 40, 80, 160, and 320 cells were microdissected from serial mouse cecal tumor tissue sections. Matrix solution was micro-deposited onto the captured cells. For MALDI MS analysis, the laser intensity remained consistent throughout the experiment and one hundred laser shots were acquired per mass spectrum. The mass spectra obtained from these cells were shown in figure 7. Protein signals can be detected with as few as 5 cells. As the number of microdissected cells increases, the signal to noise ratio and number of peaks detected increase gradually from 20 cells to 160 cells and level off between 160 and 320 cells. Therefore, capturing at least of two hundred cells was enough for obtaining high quality mass spectra, while also dramatically reduces the microdissection time and allowing for analysis of small amount of target cells.

Reproducibility

The reproducibility of this technique is evaluated with protein profiles obtained from two independently prepared tissue sections from a mouse cecal tumor sample. The tumor sample is very homogenous according to the histology and the same numbers of cells were captured by LCM and analyzed using MALDI MS. Mass spectra were acquired from five different regions and averaged to represent the protein expression profile from one tissue section. Similar experiments were performed two weeks later on the same frozen tumor. The MALDI spectra were baseline corrected and the intensities were normalized to the total ion current. The intensities of the 164 most intense peaks found in the two spectra were compared, as shown in figure 8. The results show that 151 of these 164 peaks are within 1.5-fold changes for the two

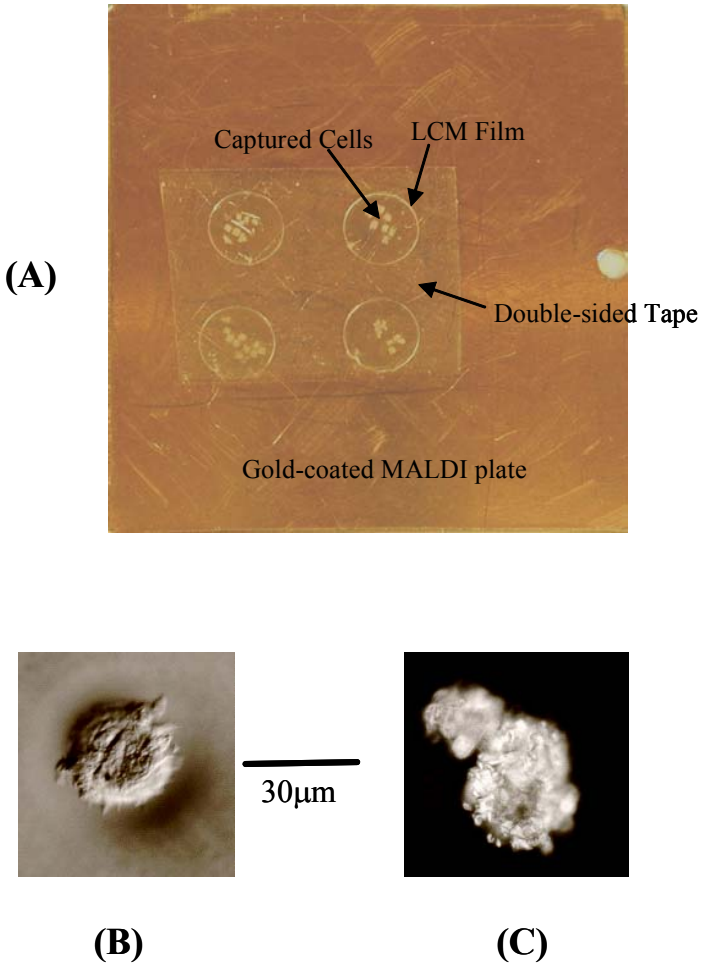


Figure 6: LCM films adhered to a MALDI target plate with double sided tape (A). (B) shows the captured cells on the LCM film prior to matrix deposition, and (C) shows the matrix solution crystallize on top of the captured cells.

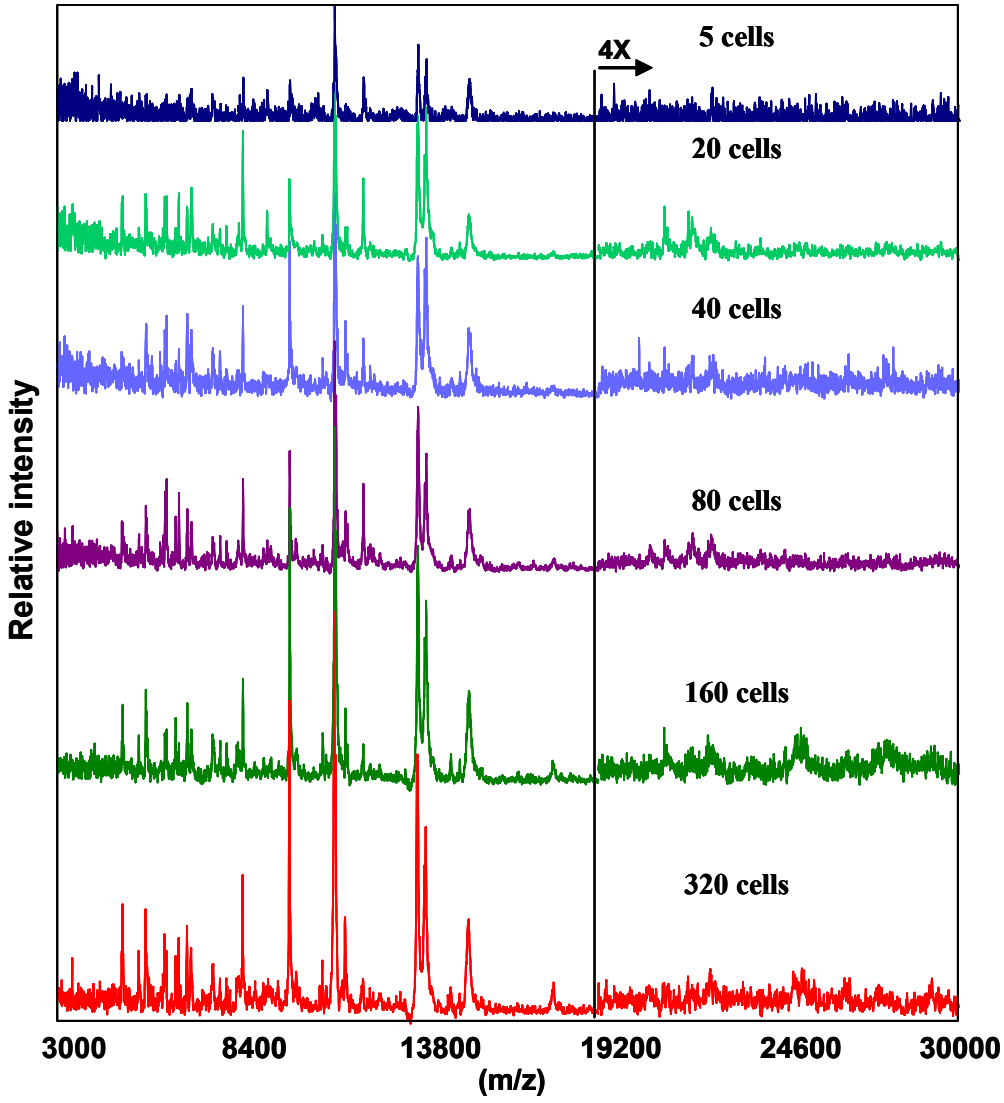


Figure 7: The sensitivity assessment for combining LCM and MALDI MS. Protein signals were observed from 5 cells and the signal-to-noise ratio were improved slightly as the number of cells captured increases.

different analyses, and no peaks were found having more than 2 fold changes. The R squared value is 0.968 for the linear regression fit. This result demonstrates that highly reproducible mass spectra are obtained using this technology.

Dehydration Effect

The dehydration procedure required by LCM contains several graded alcohol wash steps which may remove some highly soluble proteins from the tissue sections. To evaluate the effect of dehydration on obtaining reproducible protein profiles, three mouse liver sections were treated as follows: a) a control tissue section with no treatment; b) a tissue section was emerged in 70% ethanol for 1min; c) a tissue section was emerged in water for 3 minutes. Matrix solution was then deposited on the three tissue sections at similar regions and MALDI mass spectra were generated from these matrix spots. No significant differences in the three spectra were found (Figure 9), which shows that the tissue dehydration process required in preparation for LCM does not cause significant protein loss from the tissues. Though a few peptides were found by subsequent MALDI analysis of the washed ethanol solution, these same peptides were also detected from the washed tissue sections with similar intensities compared to control, suggesting sufficient proteins remained in the tissue after the wash procedure. Some physiological salts from the tissue, which interfere with the MALDI ionization process, are also washed away in the dehydration procedure. Thus washing steps can help improve the MS quality.

Comparison with Direct Tissue Protein Technique

Other sample preparation steps required by LCM, such as cell capture and transfer via the thermoplastic film, may influence the protein profiles when comparing them with the results from direct tissue analysis. To evaluate the potential differences, a mouse cecal tumor, which was homogenous according to the histology, was cut at 5 μm sections. For the LCM approach,

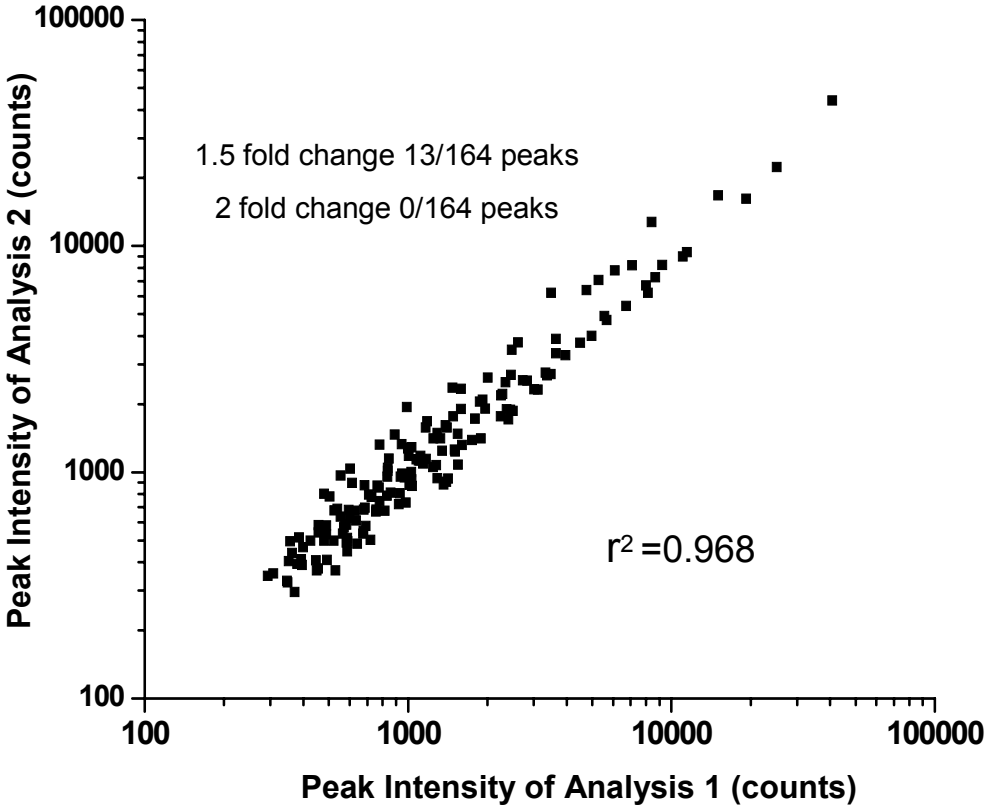


Figure 8: Reproducibility analyses of protein expression profiling with LCM combined with MALDI MS. Comparison of two spectra obtained from the same sample but different independently prepared sections at different days are presented.

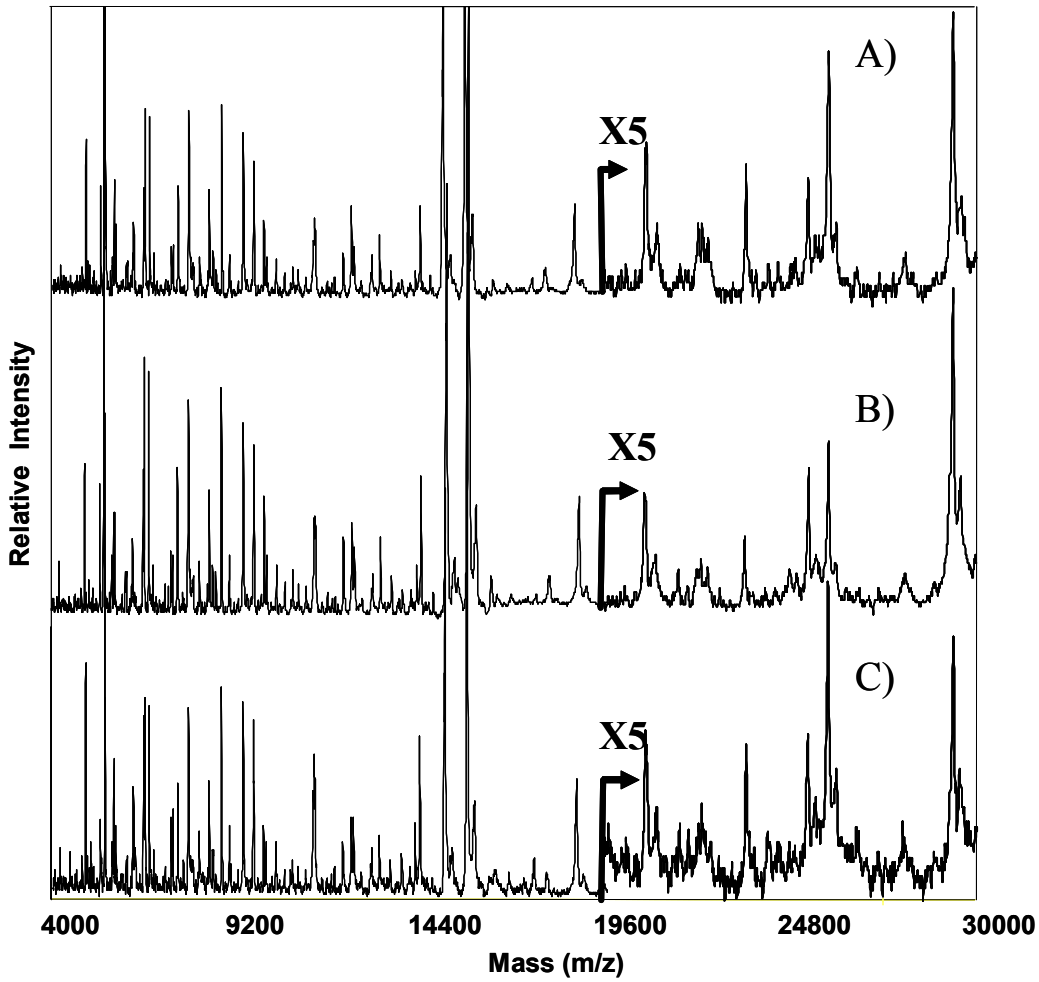


Figure 9. The tissue dehydration process required in preparation for LCM does not result in significant protein loss from the tissues. MALDI MS spectra of A) Untreated mouse liver, B) mouse liver after exposure to 70% ethanol for 1 minute, and C) mouse liver after exposure to water for 3 minutes.

approximately 2000 tumor cells were captured from the tissue sections. Seven individual mass spectra were obtained and averaged as one (Figure 10A). For the direct tissue analysis, six 150 nl droplets of matrix solution were deposited on the tissue sections. Six mass spectra were obtained from the tissue sections and averaged as one (Figure 10B). The experiment was repeated three times on three different samples and similar results were obtained. For the comparisons, the overall protein profile qualities acquired using both approaches were similar in the number of signals detected and peak mass range. Though the mass spectra obtained using the direct tissue analysis approach had better signal-to-noise ratio and resolution for some peaks. Some peak intensity differences were also observed in the two mass spectra, such as the peaks detected at 4898, 9738, 11055 and 15711 m/z. Therefore, using one consistent technique is necessary for tissue protein profiling and comparison.

Staining Effect

Histological staining facilitates the recognition of various cell types in tissue and is commonly used in pathology. Haematoxylin and eosin (H&E) is a widely used staining method and its effect on obtaining protein profile using MALDI MS was evaluated with mouse liver sections. Both the control tissue and H&E stained tissue sections went through the same dehydration preparation steps, except the control section skipped the H&E staining. Figure 11 shows the mass spectra obtained from H&E stained and controlled tissue sections. The spectrum obtained from the unstained cells (Figure 11 B) clearly shows more signals and better resolution when compared to those of stained cells (Figure 11 A). Therefore, H&E tissue staining prior to LCM severely interfered with the MALDI mass spectrum quality.

Since histological staining is necessary for recognition of some special cell types, development of a staining protocol which is compatible with MALDI MS analysis is important

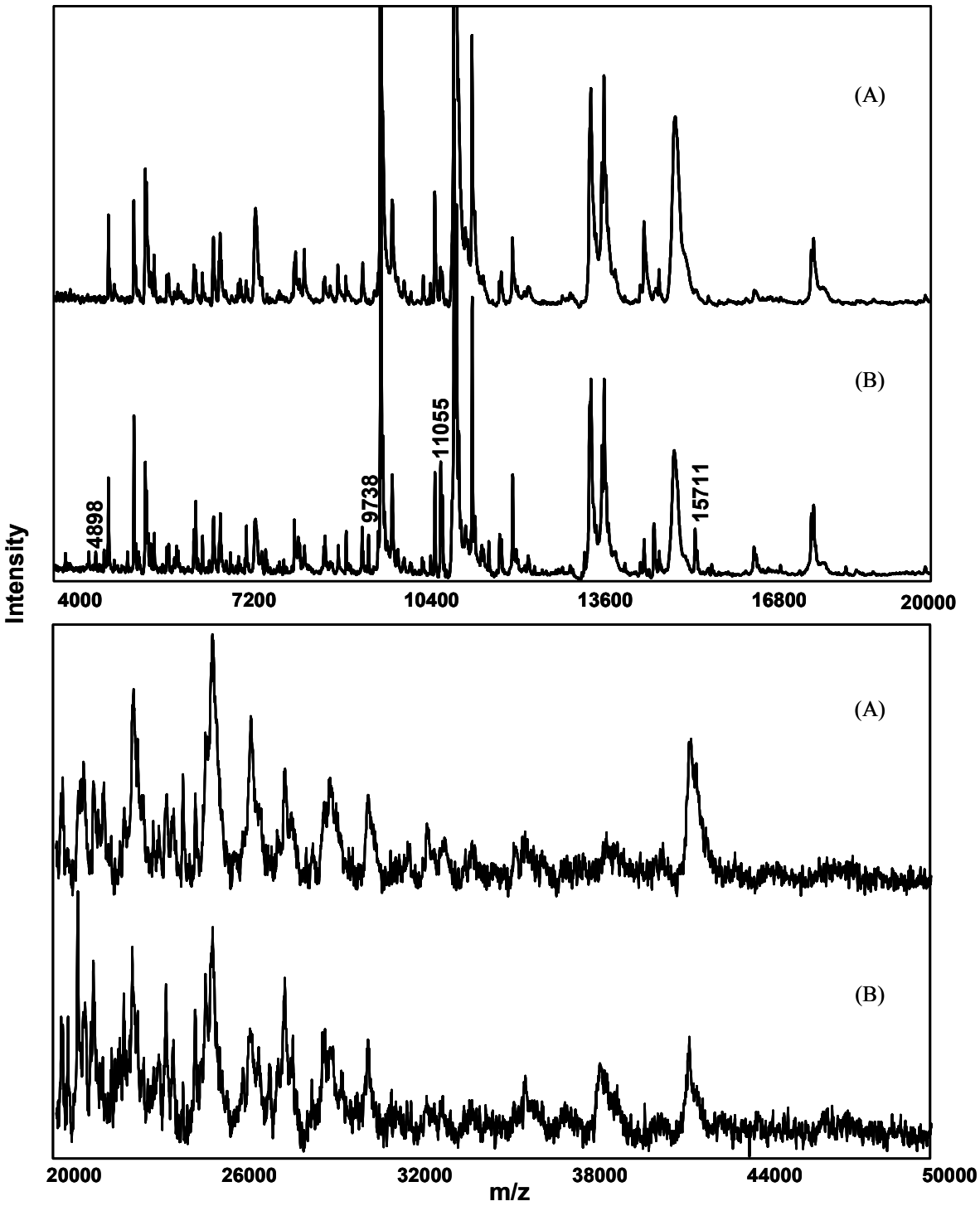


Figure 10: The comparison of MALDI MS from (A) LCM cells and (B) direct tissue sections. The overall mass spectra qualities were similar when obtained using both of the approaches, though peak intensity differences were observed for some peaks as labeled.

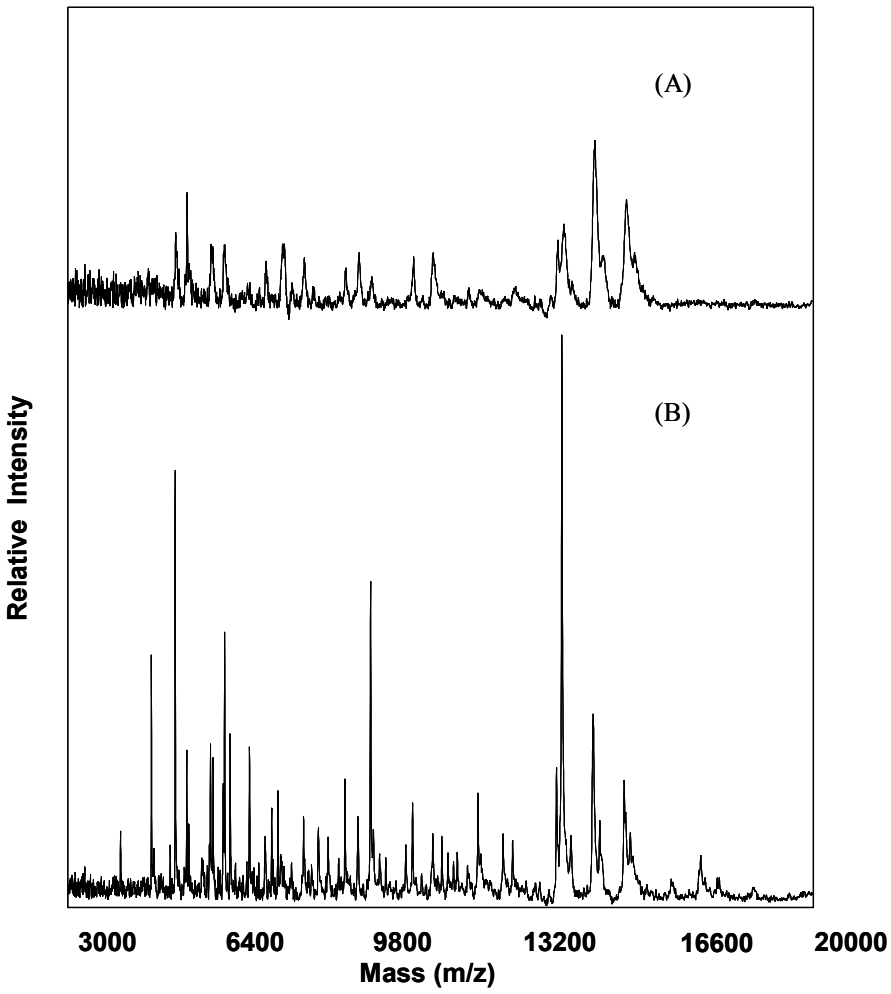


Figure 11. The histological staining of tissue prior to LCM reduces spectrum quality. (A) MALDI MS spectrum obtained from haematoxylin and eosin (H&E) stained LCM captured mouse liver cells. (B) MALDI MS spectrum obtained from unstained LCM captured mouse liver cells.

for accurately selecting cells of interest using LCM. The compatibilities of cresyl violet and methylene blue with LCM and MALDI MS analysis were further evaluated [52]. Consecutive mouse liver sections were cut and stained with cresyl violet and methylene blue. Approximately 400 hepatic cells were microdissected by LCM and analyzed using MALDI MS. For both Cresyl violet and methylene blue, the dehydration steps did not destain the sections. Figure 12 shows the protein profiles obtained from unstained, methylene blue and cresyl violet stained microdissected cells in the m/z range from 3000 to 30000. The overall spectral quality, resolution and signal intensity, were evaluated and found to be in good agreement with the signals recovered from unstained control section. The spectra obtained from methylene blue-stained cells were found remarkably similar to those obtained from control cells over the entire studied m/z range. The profiles acquired from cresyl violet-stained cells showed some signal intensity differences from the control profiles. For example, the signals at *m*/*z* 9911, 14287, and 29281, which were detected with strong intensity in the control cells, were detected with lower intensities after cresyl violet staining. As a general trend, stronger intensity signals were obtained after cresyl violet staining in the *m*/*z* range below 12000. However, signals in the *m*/*z* range above to 18000 were better detected from the control or methylene blue stained cells.

LCM Film Effect

The thermoplastic polymer film used in LCM is non-conductive and may accumulate excess charges on the surface during the MALDI process. These accumulated charges may distort the local electrical potential and interfere with the mass accuracy and peak resolution. Therefore, internal calibration is applied for each MALDI MS analysis to compensate for the surface charging effects. To evaluate the mass accuracy with internal calibration, standard proteins, insulin and apomyoglobin, were spotted on the LCM film and measured using MALDI MS. With the internal calibration of insulin and apomyoglobin peaks, the mass accuracy for the doubly

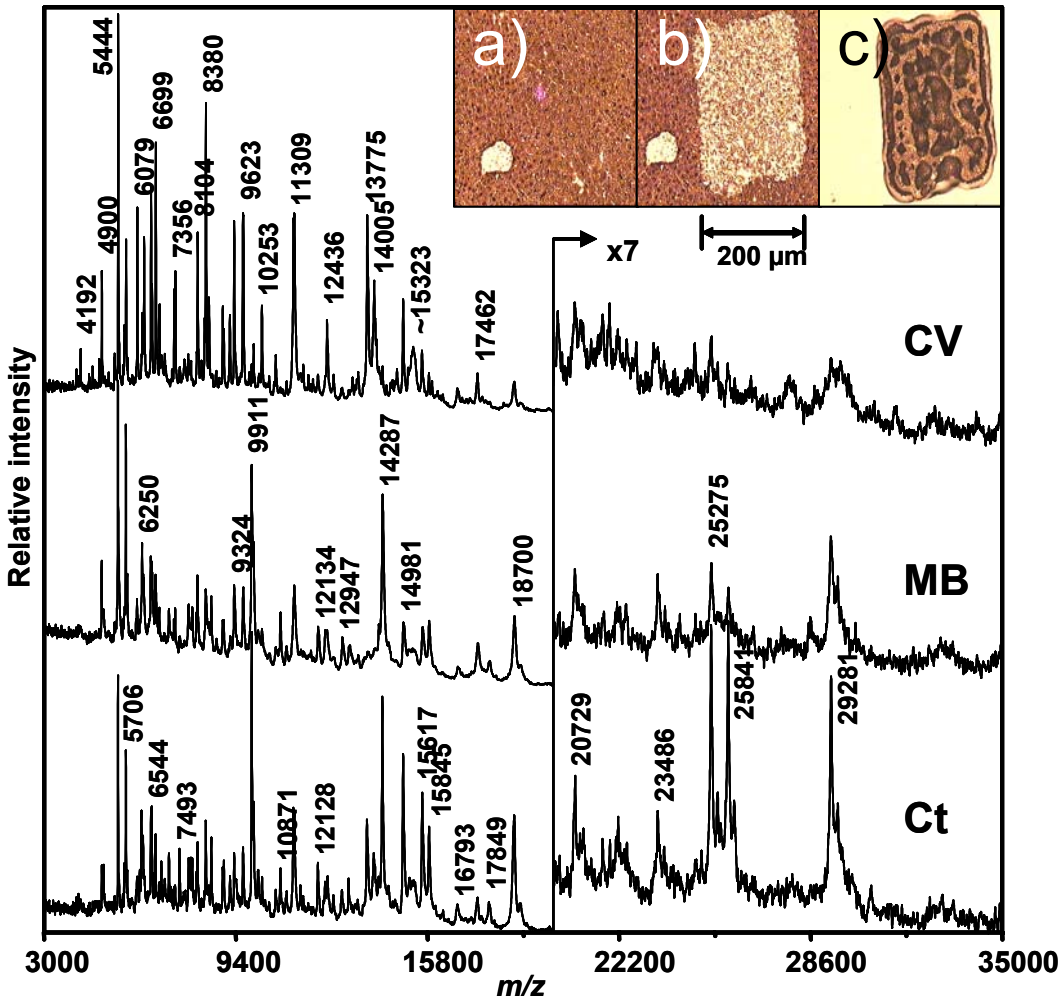


Figure 12 (a-c) LCM of Cresyl Violet-stained mouse hepatic cells. MALDI-MS protein profiles obtained from microdissected mouse hepatic cells. Ct, control; MB, stained with Methylene Blue; and CV, stained with Cresyl Violet.

charged apomyoglobin peak was evaluated. Fourteen consecutive 100 laser shot mass spectra were acquired, and an average mass accuracy of 60 ppm with 40 ppm standard deviation was obtained. This mass accuracy is acceptable for linear MALDI MS analysis. Therefore the mass accuracy did not suffer using the LCM procured cells with internal calibration.

Another possible effect caused by the polymer film is that polymeric peaks may be observed in the lower mass range if the MALDI laser intensity dramatically exceeds the normally required level. Polymer peaks can be detected from the LCM film and are usually under 4000 Da (figure 13). Abundant protein signals across a wide mass range can be detected without generating these polymer peaks with proper adjusted laser intensity (figure 7, 8 and 12). Therefore the MALDI laser intensity should be controlled properly to avoid possible polymer peaks.

Summary and Conclusions

With the optimized sample preparation procedure, LCM combined with MALDI MS is a sensitive, accurate, fast and reliable method to acquire protein profiles from a specific cell population sampled in a heterogeneous tissue. In this chapter, the sample preparation protocol for providing high quality mass spectra is described and various aspects of this technology are evaluated. The matrix solution microspotting technique and development of MS compatible histological staining significantly improved the reliability and sensitivity of this technology. The dehydration process and the LCM film were evaluated and did not show significant interference with the protein profile quality.

LCM combined with MALDI MS has great applications in a wide range of biological systems. Its unique ability to quickly and accurately generate protein profiles from small numbers of pure cells from a heterogeneous tissue has extraordinary utility in finding disease marker, elucidating pathomechanisms, and classifying tissue subtypes at the protein levels. It also has the

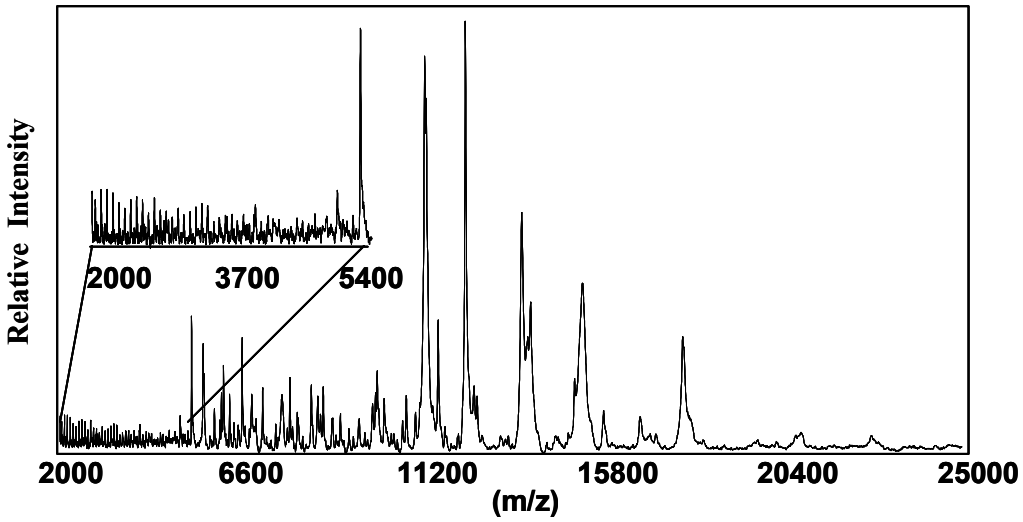


Figure 13: The polymer peaks are found on from LCM film when MALDI laser intensity is too high. The polymer peaks presents under 4000 Da and can be eliminated by lowering the MALDI laser intensity.

potential to play an important role in clinical applications for disease diagnosis and prognosis.

Materials and Methods

The sinapinic acid, cresyl violet, methylene blue and the standard proteins were purchased from Sigma Chemical (St. Louis, MO). Trifluoroacetic acid (TFA) was from Burdick and Jackson (Muskegon, MI). HPLC grade acetonitrile was from EM Science (Merck, Darmstadt, Germany). The ethanol was from AAPER alcohol, absolute 200 proof (Shelbyville, KY) and xylene is from Pharmco (Brookfield, CT). The frozen mouse cecal tumor samples are a courtesy from Dr. Nipun Merchant.

Tissue sections were cut on a CM1800 cryostat from Leica Microsystems (Wetzlar, Germany). Laser capture microdissection was performed using the Arcturus PixCell II LCM system and the CapSure caps were from Arcturus Co. (Mountain View, CA). For LCM, the laser duration time was 800 microseconds and the laser power was 75mW. The double-sided tape is from Digi-Key Co (Thief River Falls, MN).

The 7" length Drummond glass capillary used for oocyte injection from Drummond Scientific Co. (Broomall, PA, USA) is utilized for making the matrix solution deposition capillary. The capillary is pulled in a P-87 Laser Based Micropipette Puller from Sutter Instrument Co (Novato, CA, USA). The pulling program is set as follows: First cycle: Heat 780, Pull 50, Vel 26, Time 35; Second cycle: Heat 780, Pull 50, Vel 10, Time 15. After the glass capillary is pulled, the opening of the tip is cut using a razor blade, leaving the opening of the capillary approximately 10 micrometer in diameter.

Methylene blue is prepared in 0.15% in 70% ethanol. The solution is stirred overnight and filtered before use. Cresyl violet was prepared in 0.5% deionized water and stirred overnight on low heat. The cresyl violet solution was filtered and 2 drops of glacial acetic acid was added to a 100-mL solution. The tissues were stained 10 seconds in methylene blue and 30 seconds in

cresyl violet solution, and then followed with the graded ethanol and xylene dehydration process. Hematoxylin and eosin were from Richard-Allan Scientific (Kalamazoo, MI). The hematoxylin and eosin stains procedure was 1 min in 70% ethanol, 30 s each in hematoxylin followed by eosin, 20 s in water, followed by previously described ethanol and xylene wash steps.

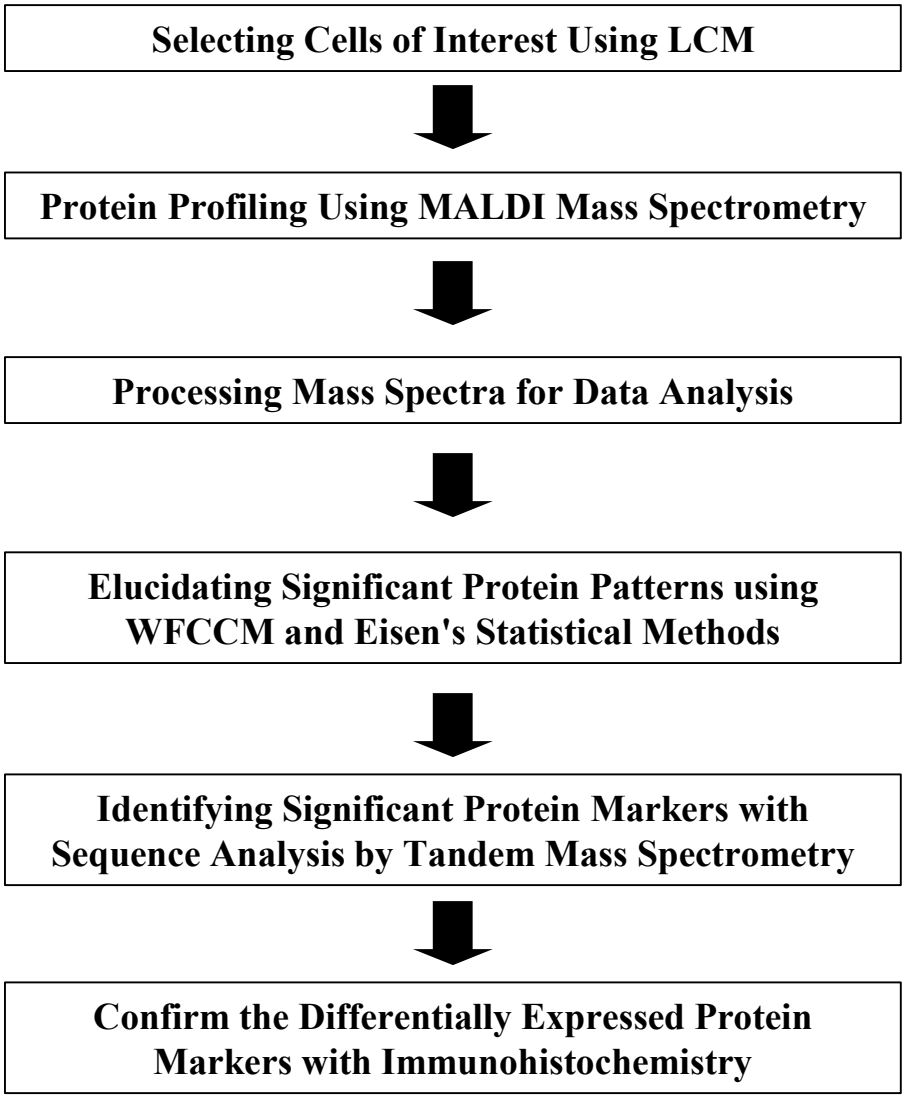
MALDI-MS analysis was performed using a Voyager DE-STR MALDI-TOF mass spectrometer from Applied Biosystems (Framingham, MA) with a 337 nm nitrogen laser. The mass spectra are acquired in the linear mode to accommodate the wide mass range. The optimized delayed extraction condition for this research is set as follows: the accelerating voltage is 25,000 V, the grid voltage is 91%, and the delay time is 220 nano-seconds. To obtain the best sensitivity, the guide wire is set at 0.055%.

APPLICATIONS TO BIOLOGICAL SYSTEMS

Introduction

The optimized protocol developed for direct protein profiling from laser microdissected cells using MALDI MS was described in chapter II. Many significant biological questions which cannot be easily studied with other proteomic technologies can be addressed using this technology. This technique can be virtually applied to any type of frozen tissue. Several different applications are discussed in this chapter, including finding early protein markers in normal and neoplastic mouse colon, classifying human breast tumors subsets, and predicting rat kidney glomerulosclerosis. The general strategy for these studies is shown in Table 2. Briefly, the specific populations of cells were first obtained from the frozen tissue sections using laser capture microdissection. Second, the protein expression profiles were acquired using MALDI mass spectrometry. Third, after proper data processing, such as mass spectra calibration, baseline correction and peak binning, multi-functional statistical analyses were performed and statistically significant protein markers were selected. Proteomic patterns unique to each tissue group were obtained. The different tissue groups were clustered using these proteomic patterns. Finally, the most significant protein markers were separated from the tissue lysate and identified through amino acid sequence analyses using tandem mass spectrometry. Further functional analyses of these differentially expressed protein markers will not only advance our understanding of the pathomechanisms and aid in elucidating potential drug targets, but the disease specific proteomic patterns also have the potential to serve as diagnostic and prognostic tools for better patient care and treatment.

Table 2. Overview of the general strategy for elucidating differentially expressed protein markers using LCM combined with MALDI MS technology.



CHAPTER III

PROTEIN PROFILING OF NORMAL AND NEOPLASTIC MOUSE INTESTINE

Introduction

Colorectal cancer (CRC) is the fourth most common cancer and the second most common cause of cancer deaths in the US, with estimates of more than 146,940 new cases and 56,730 deaths for the year 2004 [53]. Although significant improvements have been made in the treatment of CRC with surgery, chemotherapy and the combination of radiotherapy-chemotherapy, patients with advanced CRC still have a poor prognosis. Early diagnosis and improved understanding of the biological behavior of each tumor is vital for the success of treatment. Identification of early colorectal tumor markers at the protein levels is expected to make significant improvements for better understanding of CRC mechanisms, as well as assisting clinical diagnosis, prognosis and treatment for CRC patients.

Colorectal neoplasia initially arises from colonic crypts (Figure 14A). A normal mouse crypt is only 30 micrometers in width and 250 micrometers in length (Figure 14B). Crypt cells continue to rapidly proliferate throughout life. A small number of stem cells situated near the bottom of the crypts give rise to various daughter cells, such as absorptive cells, endocrine cells, goblet cells, and tuft cells [54-56] (Figure 15A). These daughter cells migrate from the crypt base to the luminal surface and are shed into the lumen of the colon. Steady supplies of cells are generated in the bottom crypt compartment to replace the shed cells (Figure 15B). In mouse, daughter cells migrate from the bottom crypt to the top in approximately 7 days with up to 60% of the crypt cells dividing twice daily [57]. Hence the differentiated cells are found mainly toward the top of the colonic crypts. Currently, little is known about the factors that control the stem cells proliferation, differentiation and their roles during malignant process.

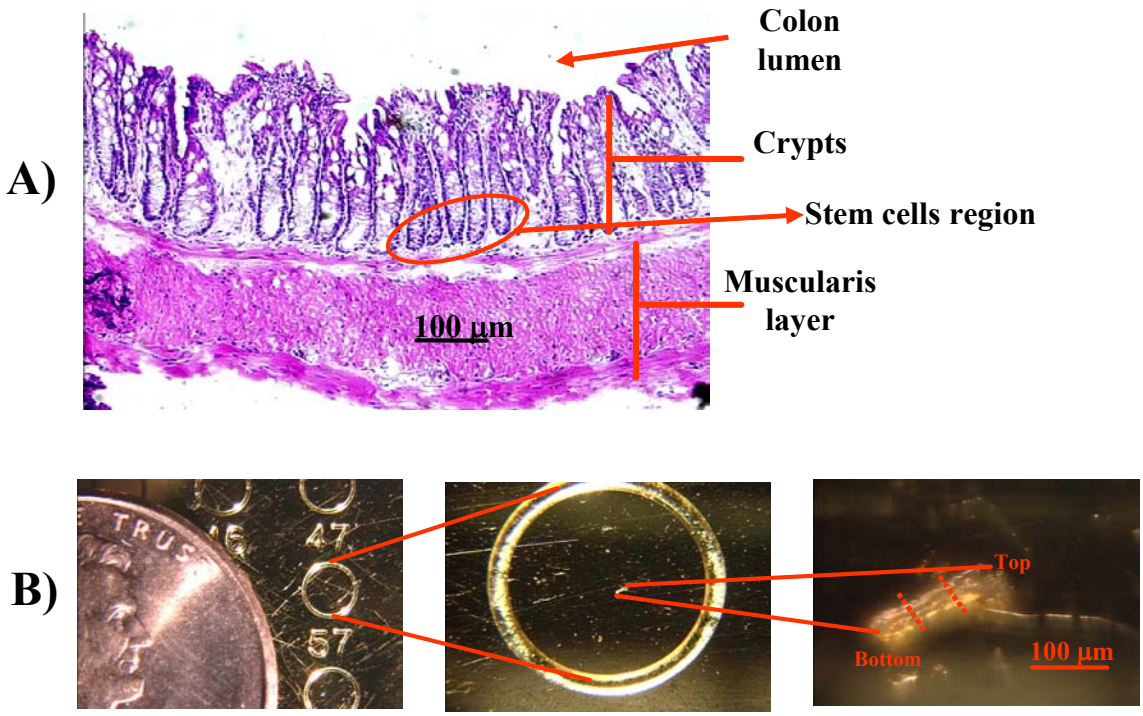


Figure 14: The histology and size of mouse colonic crypts and polyps. A). The histology of a mouse colon with normal crypts. The top compartments of crypts open to the lumen of the gut and stem cells reside at the crypt base. B) An isolated crypt on a MALDI plate with a penny for size reference. The average length for mouse colonic crypts is 250 µm. The top and bottom compartments were separated with dotted lines. C). The histology of adenomas from an *Apc*$^{Min/+}$ mouse polyp.

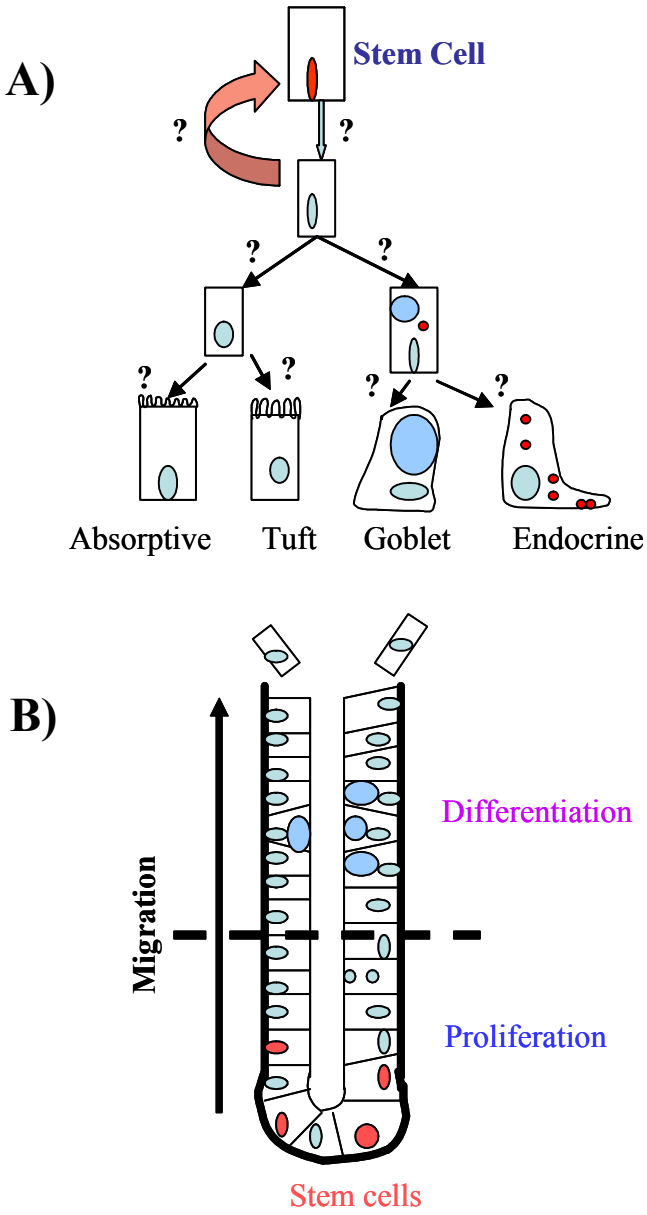


Figure 15: The biology and structure of colonic crypt. A). The stem cells give rise to various daughter cells. Little is known about the factors that control the stem cell proliferation and differentiation. B), The cells situated at the crypt base constantly proliferate and migrate up to the top crypts and shed away into the lumen.

The proliferative, progenitor compartment is speculated to have a different protein expression profile compared to the post-replicative, differentiated compartment in the upper portion of the crypt. Stem cells at the crypt base cannot be identified morphologically or distinguished using molecular markers from their surrounding epithelial cells. Identifying the unique protein patterns for the top and bottom crypt compartments may advance our understanding for the mechanisms that control stem cells renewal, proliferation and the roles they play in colorectal neoplasia development. Therefore, the proteomic patterns from the normal top and bottom of crypt compartments serve as important references for comparison with colonic neoplasia.

During colorectal cancer development, adenomas or polyps are the early benign neoplasia originated from crypts. An excellent animal model for obtaining adenomas in intestines is the *Apc*$^{Min+/-}$ mouse model. The multiple intestinal neoplasia (*Min*) mouse model is a germline mutant mouse model of gastrointestinal malignancy [58] with a mutation in the adenomatous polyposis coli (*APC*) gene. The *APC* gene codes for a 312 kDa cytoplasmic protein with 2843 amino acids. Loss of *APC* gene function is often the first genetic event in the cascade of genetic events that lead to malignant transformation in colon. Mutations in the *APC* tumor-suppressor gene have been detected in ~80% of all human CRC [59]. Alterations in the *APC* gene prevent proteasome-mediated degradation of β-catenin, a key player of the Wnt signaling pathway. As a consequence, β-catenin is stabilized in the cytosol and translocated to the nucleus, where it participates in the transcription of a number of Wnt responsive genes and initiate colonic neoplasia transformation [60]. Intercellular adherens junctions and cell contacts regulation were also increased as result of elevated level of E-cadherin from *APC* gene mutation [61]. These changes in cellular contacts may explain how *APC*-mutated cells manage to remain within the crypts. In addition, *APC* is an important regulator of the microtubule cytoskeleton and plays a role in cell migration [62]. *Apc*$^{Min/+}$ mice develop on average of 35 adenomas in the small intestine and colon [63]. Similarly,

familial adenomatous polyposis (FAP) human CRC patients have mutations in the *APC* gene and develop hundreds of adenomatous polyps in colon and rectum.

The limited size of mouse colon crypt presents a challenge for conventional proteomic techniques such as 2-D gel electrophoresis. Different regions of each crypt only contain around 10 cells in 10 μm thick tissue sections. Thus, the analysis of different crypt compartments requires a special technique to harvest these very limited cell populations. The adenomas from $Apc^{Min+/-}$ polyps are heterogeneous, composed of epithelium, stroma and various other cells (figure 16). The tumor cells should be specifically selected to compare with epithelial cells from normal crypts. Laser capture microdissection combined with MALDI-TOF MS offers a powerful technique to analyze the protein expression profiles from these limited and specific cells of interest. In this study, the protein profile analyses were performed using LCM and MALDI MS. Differentially expressed proteomic patterns were found that upon statistical analyses distinguished normal crypts from adenomas. Significant protein markers were identified through amino acid sequence analysis using tandem mass spectrometry. Immunohistochemistry was performed on some of the protein markers to confirm their differential expression.

Specific Aim

To identify specific proteomic patterns unique to different colonic crypt compartments and $Apc^{Min+/-}$ adenomas to aid early colorectal cancer diagnosis.

Results

Protein Profiles of Top and Bottom of Normal Crypts and Adenomas

Normal crypts were microdissected from adult wild-type C57/BL6 mice. The LCM process is illustrated in Figure 17. Different crypt compartments are clearly recognized from the

Figure 16. The heterogeneous nature of *Apc*Min+/- adenoma with adjacent normal colon.

longitudinal cut crypts (Fig 17a). In each tissue section, the bottom third of the crypts was first microdissected using LCM (Fig 17b). The cells from the top of the crypt were also microdissected using LCM for comparative purposes (Fig 17c). Although a single LCM laser shot captures approximately ten crypt cells (Fig 17d), which was sufficient to generate a good quality mass spectrum, we chose to capture approximately 10 crypt bottoms to obtain a more broadly representative cell population. Capturing more than 100 cells did not significantly improve the quality of the mass spectra. The microdissected cells from the top and bottom of crypts are shown in Figure 17e and 17f, respectively. A total of 68 and 62 individual groups of cells were captured from 4 adult wild-type C57/BL6 mice. Similarly, epithelial cells of adenomas from *Apc*$^{Min+/-}$ mice polyps were obtained using LCM with minimal amount of the contamination of surrounding stroma cells. Approximately 100 cells were captured for each group of microdissected adenomas. A total of 78 individual groups of adenomatous cells were microdissected from 8 adult C57/BL6 *Apc*$^{Min+/-}$ mice.

The protein expression profiles were directly acquired from the microdissected cells using MALDI-TOF MS. Each mass spectrum was acquired from an individual group of microdissected cells. The representative mass spectra for the top and bottom of the crypt and adenomas are shown in figure 18A-C. Different protein expressions were shown in the three groups. For each mass spectrum, approximately 200 significant signals were detected in the protein mass range 2000 to 70,000 Da, with proteins under 20,000 Da yielding the best resolution. After internal calibration, the spectra were baseline corrected and normalized using an algorithm developed in our laboratory.

Proteomic Patterns Classifications among the Top and Bottom of Normal Crypts and Adenomas

A total of 68 and 62 individual MALDI mass spectra were obtained from the normal top and bottom crypts, respectively and 78 mass spectra were obtained from the adenomas. The

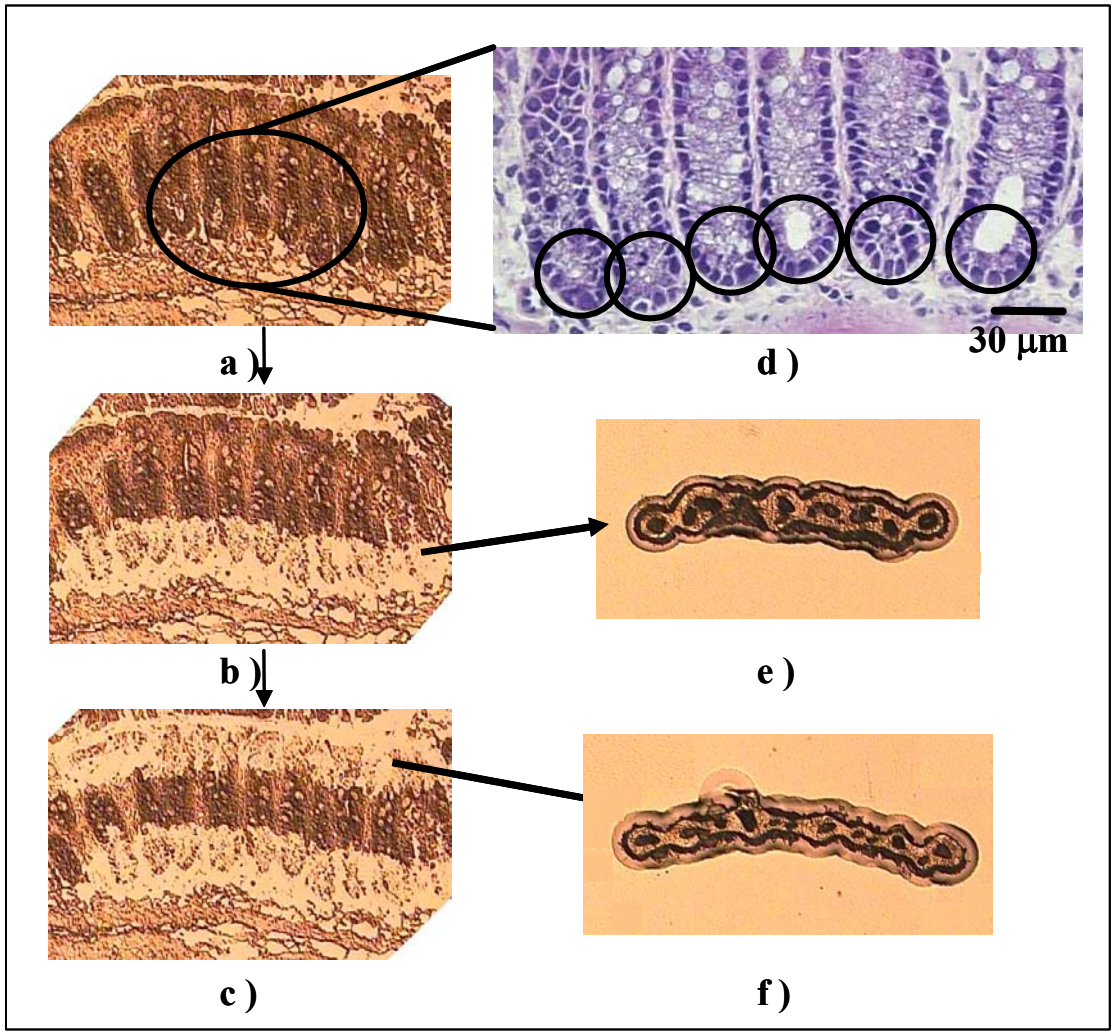


Figure 17: The process of using LCM to microdissect the specific cells from the top and bottom of crypts. a) longitudinal crypts were chosen for microdissection; b) and c) shows the crypts after the bottom and top compartments were microdissected, d) To illustrate the size and positions where LCM laser hits, black circles were shown on the parallel cut tissue section stained with H&E. Approximately 10 cells were captured per laser hit. The scale bar is 30 μm; e) and f) show the captured cells from the bottom and top crypt compartments on the LCM films.

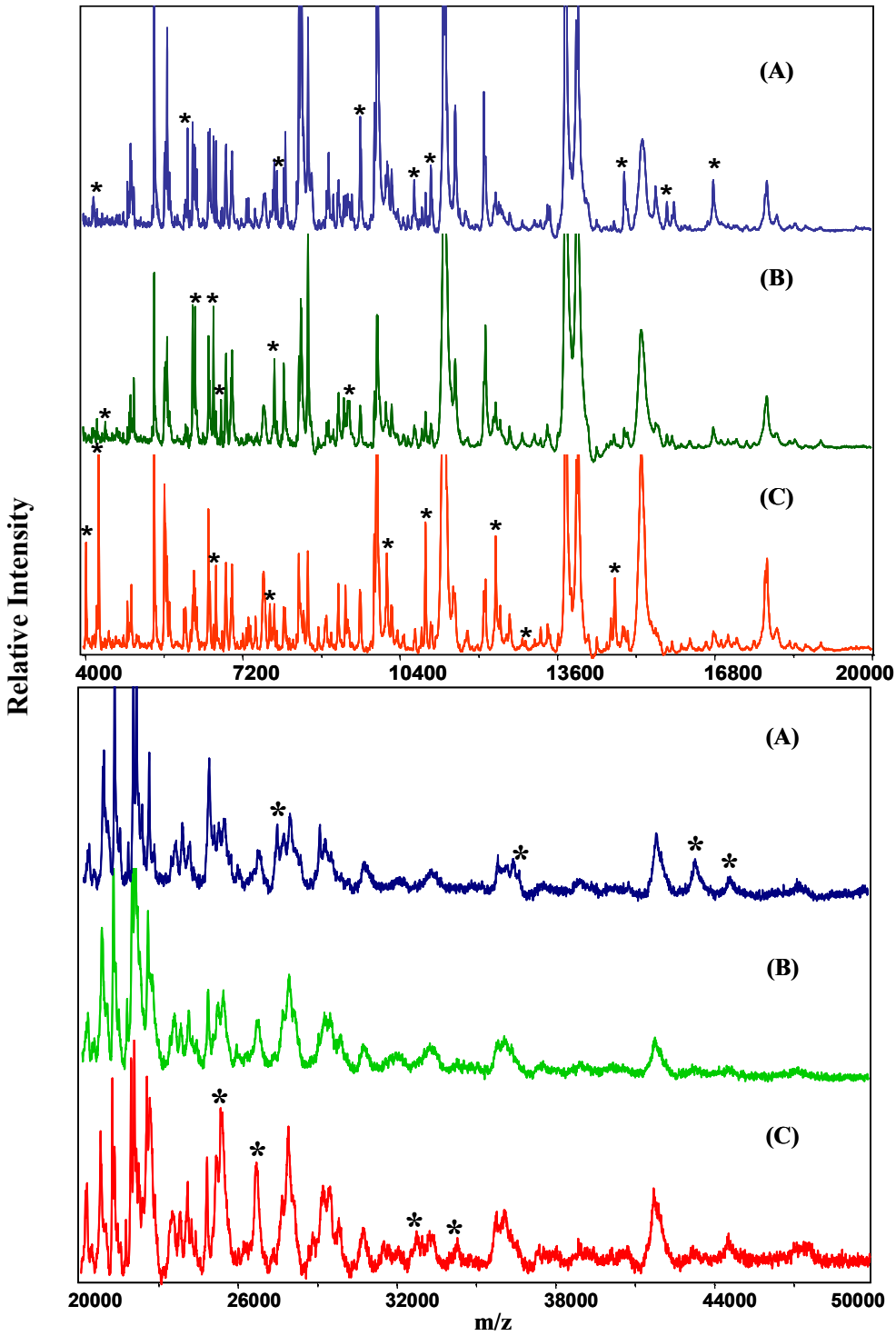


Figure 18: The different protein expression profiles for the top (A), bottom murine colonic crypts (B) and murine colonic adenomas (C). As marked with a *, different protein expression patterns were observed in the different intensities in the mass spectra.

signals across all these spectra were compared and over 1400 distinct peaks were obtained. To detect significant protein markers that were expressed differently among the top and bottom of crypts and adenomas, advanced statistical analysis was performed to analyze the enormous amount of data. Weighted flexible compound covariate method (WFCCM) program utilizes six different statistical tests to determine discriminatory markers [64]. This statistical method has been successfully applied to classifications of different tumor subsets in non-small cell lung cancer based on proteomic patterns [22]. According to the statistical criteria, the top biomarkers determined from the comparisons of three different groups (normal top, bottom crypts and adenomas) were obtained. The most significantly differentially expressed protein signals are shown in table 3, organized according to the classification group in which the marker was overexpressed.

Using the statistically significant biomarkers resulting from the WFCCM analyses, the protein profiles from the three different groups were compared and classified. The classification accuracy was assessed using the leave-one-out cross-validation class prediction method (Table 4). The protein profiles from the normal top and bottom crypts were successfully distinguished from each other with 97% accuracy based on the top 77 differentially expressed protein markers. This result clearly demonstrates the proteomic differences that were detected between the top and bottom crypts. For the comparisons among the adenomas and different crypt compartments, 100% classification accuracy was achieved for both the top and bottom crypt compartments vs. adenomas, using the top 40 and 20 differentially expressed protein markers, respectively. Therefore, these different tissue populations are all successfully classified from each other based on the proteomic patterns.

An agglomerative hierarchical clustering algorithm [65] was used to further verify and visualize the differential protein expression patterns selected by the WFCCM analysis. For the normal top and bottom crypt comparison, 128 out of 130 protein profiles were correctly clustered

Table 3: Top differentially expressed protein markers determined by weighted flexible compound covariant method (WFCCM) analysis. The average m/z value is reported. Biomarkers are grouped according to the classification in which the marker was overexpressed.

Adenomas / Bottom Crypt	
m/z	Tissue
4060.6	Adenomas
4078.4	Adenomas
4274.1	Adenomas
4315.1	Adenomas
4442.4	Adenomas
10166.9	Adenomas
14791.8	Adenomas
14950.5	Adenomas
3326.3	Bottom Crypt
3950.2	Bottom Crypt
4450.2	Bottom Crypt
5028.9	Bottom Crypt
6799.2	Bottom Crypt
8418.9	Bottom Crypt
8432.0	Bottom Crypt
8585.8	Bottom Crypt
8982.2	Bottom Crypt
9977.8	Bottom Crypt
10718.3	Bottom Crypt

Adenomas / Top Crypt	
m/z	Tissue
4060.6	Adenomas
4078.4	Adenomas
4274.1	Adenomas
4315.1	Adenomas
5351.2	Adenomas
12373.1	Adenomas
12464.6	Adenomas
14791.8	Adenomas
14950.5	Adenomas
3326.3	Top Crypt
3950.2	Top Crypt
5465.1	Top Crypt
6120.6	Top Crypt
7491.2	Top Crypt
7855.6	Top Crypt
7934.0	Top Crypt
8028.1	Top Crypt
8338.1	Top Crypt
8418.9	Top Crypt
8432.0	Top Crypt
8492.8	Top Crypt
8542.1	Top Crypt
8626.2	Top Crypt
8639.6	Top Crypt
8982.2	Top Crypt
9078.3	Top Crypt
9452.8	Top Crypt
9977.8	Top Crypt
10184.6	Top Crypt
10589.7	Top Crypt
10718.3	Top Crypt
10827.2	Top Crypt
11055.8	Top Crypt
12435.8	Top Crypt
13471.6	Top Crypt
14983.0	Top Crypt
15619.4	Top Crypt
15852.4	Top Crypt
15995.7	Top Crypt
21179.7	Top Crypt

Bottom Crypt / Top Crypt	
m/z	Tissue
4267.4	Top Crypt
4981.3	Top Crypt
6120.6	Top Crypt
7855.6	Top Crypt
7934.0	Top Crypt
7996.9	Top Crypt
8338.1	Top Crypt
8492.8	Top Crypt
8626.2	Top Crypt
8639.6	Top Crypt
8655.5	Top Crypt
9078.3	Top Crypt
9452.8	Top Crypt
10166.9	Top Crypt
10184.6	Top Crypt
10200.5	Top Crypt
10387.5	Top Crypt
10827.2	Top Crypt
12132.7	Top Crypt
12435.8	Top Crypt
13471.6	Top Crypt
15852.4	Top Crypt
15995.7	Top Crypt
16792.9	Top Crypt
21179.7	Top Crypt
3298.8	Bottom Crypt
3326.3	Bottom Crypt
3950.2	Bottom Crypt
4004.0	Bottom Crypt
4450.2	Bottom Crypt
5028.9	Bottom Crypt
6081.3	Bottom Crypt
6799.2	Bottom Crypt
9403.6	Bottom Crypt
12654.2	Bottom Crypt
15267.8	Bottom Crypt

Table 4: Proteomic comparisons of *Apc*$^{Min/+}$ adenoma and different compartments of normal crypts. The WFCCM was used to select the statistically significant peaks that allowed the classification of the three groups of samples. The classification accuracies were calculated using the leave-one-out cross-validation class prediction method.

Comparisons	**No. of differentially expressed peaks**	**classification accuracy**
Adenoma vs. Bottom of Normal Crypt	**19**	**100%**
Adenoma vs. Top of Normal Crypt	**40**	**100%**
Top vs. Bottom of Normal Crypt	**77**	**97%**

into two groups using the top 77 differentially expressed protein markers, achieving 99% classification accuracy (figure 19A). For the comparisons of the bottom of normal crypts and adenomas, 62 bottom of crypt protein profiles and 78 adenomas protein profiles were clustered into two different groups with 100% accuracy using the top 19 protein markers (figure 19B). Similarly, 68 top of crypt protein profiles were clustered with 100% accuracy from the 78 adenomas protein profiles using the top 40 protein markers (figure 19C). All these analyses were performed as training cohorts. The red intensity scale bar corresponds to higher protein marker expression level and black corresponds to lower protein expression level. Different protein expression patterns were clearly shown in the figure 19A-C. The high classification accuracies achieved from this method further confirm the validity of the protein markers that were differentially expressed among the three groups.

Identification of Differentially Expressed Protein Markers

To further explore the biological functions of the differentially expressed protein markers, the identifications of the selected protein markers were pursued. The schematic representation for the protein marker identification strategy is illustrated in figure 20. Briefly, the normal colonic crypt and adenomas were homogenized separately and different cell lysates were fractionated using reverse phased HPLC. Each HPLC fraction was measured using MALDI MS and the fractions containing the proteins with target m/z value were chosen for trypsin digestion. The subsequent amino acid sequence analyses of the tryptic peptides were performed using capillary LC-MS/MS or nano-ESI MS/MS. The SEQUEST algorithms are used to determine the "best fit" between the MS/MS sequence and the information from protein databases [27]. With matched amino acid sequences from the MS/MS analyses, the identified protein markers have theoretical molecular weights matching the target protein molecular weights considering the post-translational modifications.

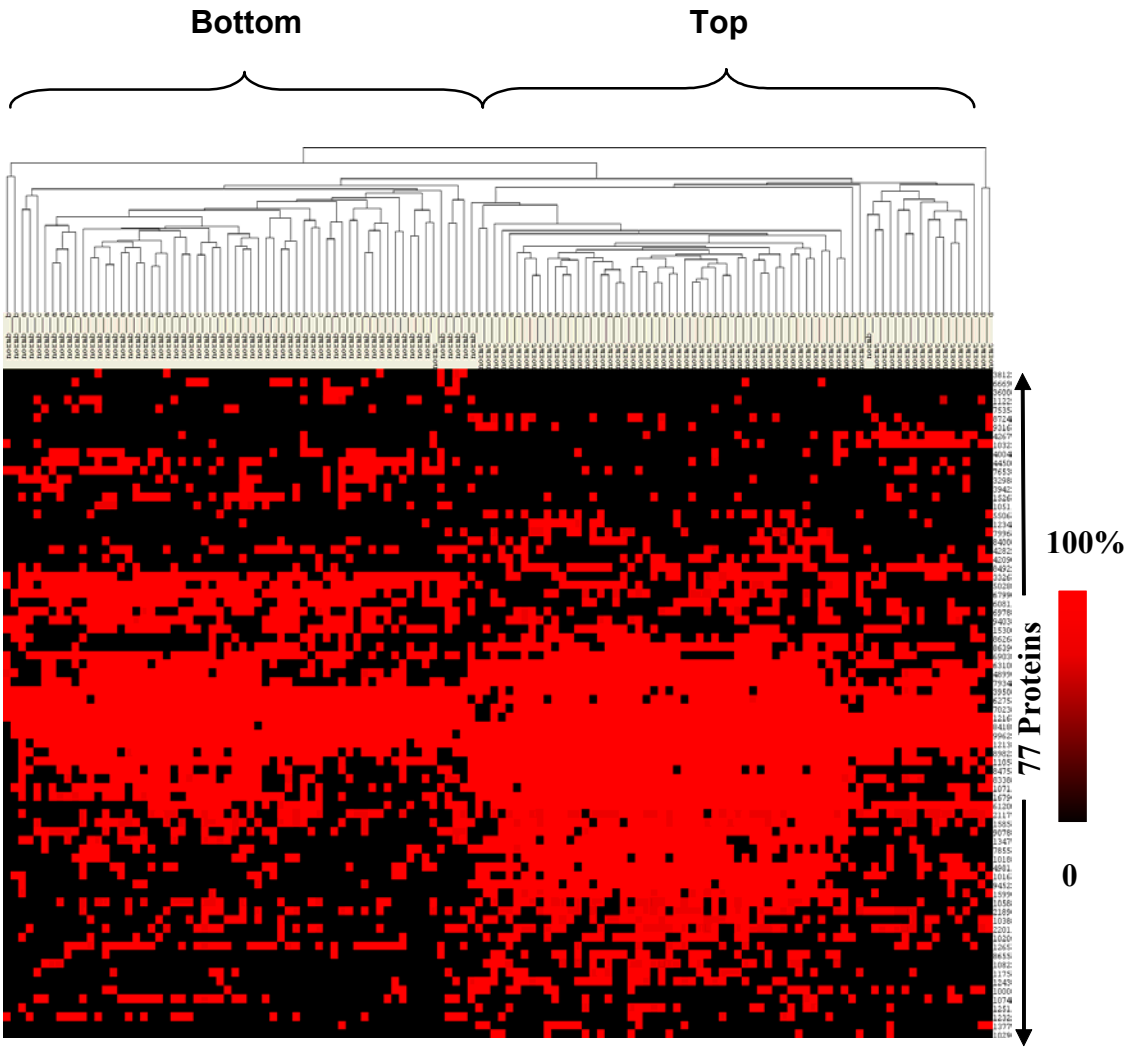


Figure 19A: Hierarchical cluster analysis of the top and bottom of normal crypts. Total of 128 out of 130 samples were classified correctly using the Eisen's clustering algorithm with the top 77 differentially expressed protein markers.

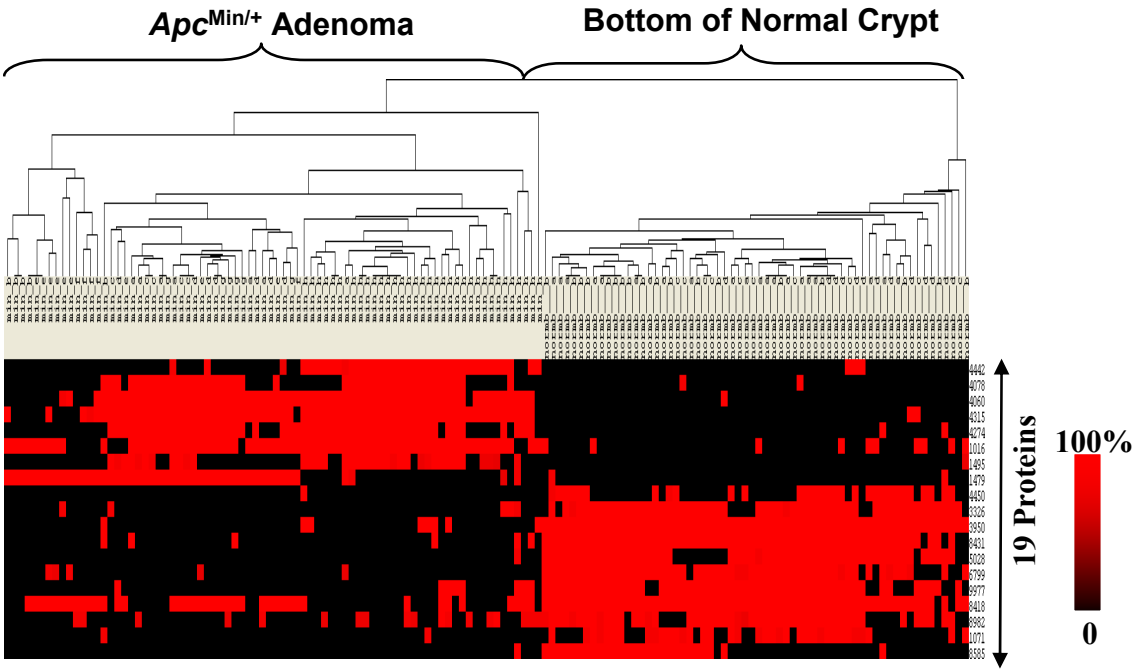


Figure 19B: Hierarchical cluster analysis of the *Apc*$^{Min/+}$ adenoma and bottom of normal crypts. A 100% classification accuracy was achieved using Eisen's clustering algorithm with the top 19 differentially expressed protein markers.

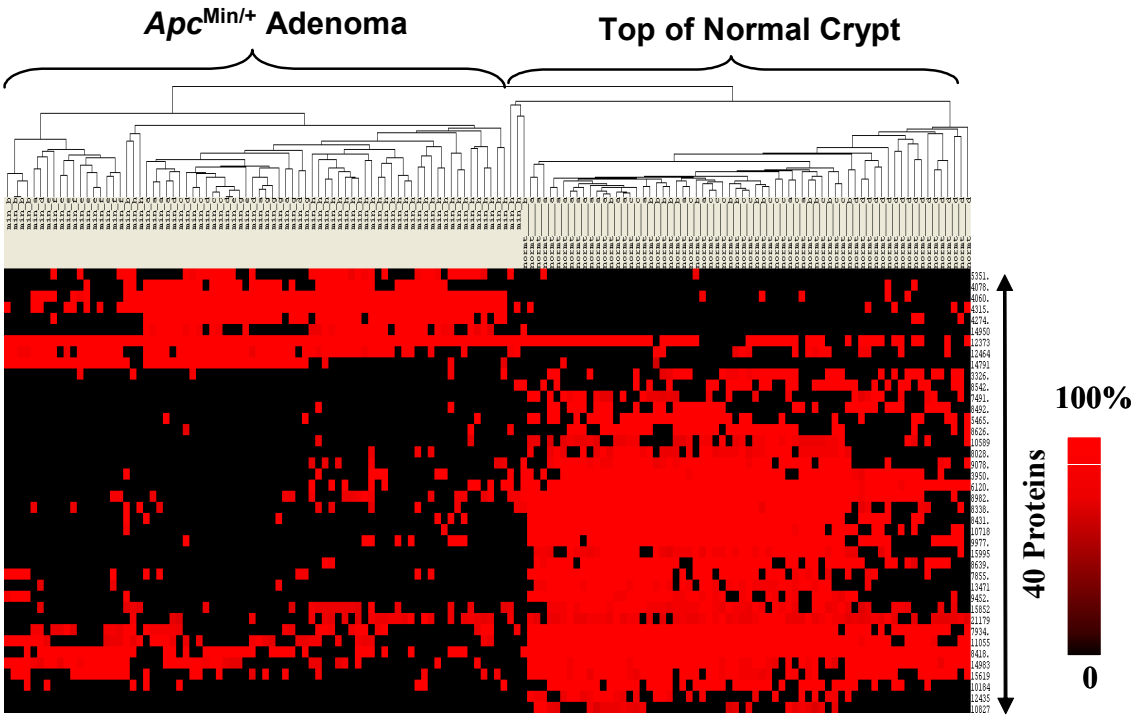


Figure 19C: Hierarchical cluster analysis of the *Apc*$^{Min/+}$ adenoma and top of normal crypts. A 100% classification accuracy was achieved using the Eisen's clustering algorithm with the top 40 differentially expressed protein markers.

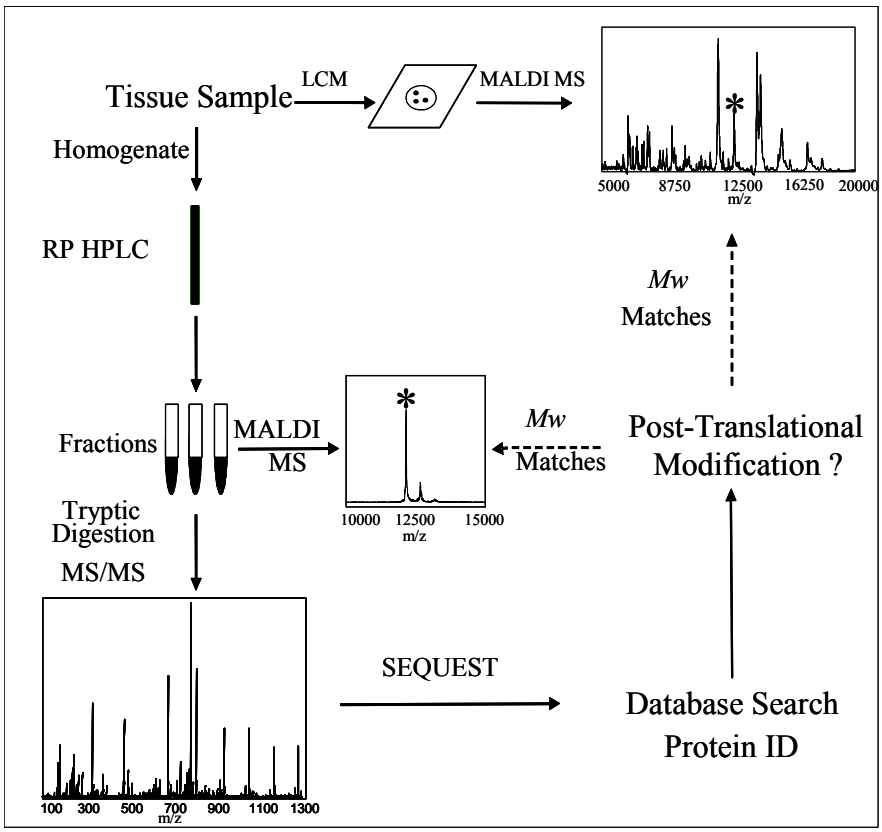


Figure 20. Schematic representation of protein marker identification process.

One example of the protein identification is shown in figure 21. Figure 21A shows the HPLC UV (214 nm) trace for the cell lysate prepared from colonic polyps. MALDI MS indicated that fraction 26 contained the target tumor-specific proteins (Fig. 21B). The LC-MS/MS sequencing of the peptides obtained from trypsin digestion of fraction 26 provided clear evidence for the presence of galectin-2, with 4 peptides found constituting 39% of the protein sequence. Figure 21C shows the observed MS/MS sequence spectra for the peptide [ETLNLHFNPR], a unique tryptic peptide to galectin-2. Considering the plausible addition of an acetyl group at the amino-terminus after removal of the initial methionine, the average molecular weight for galectin-2 is 14790.8 Da, which matches well with the observed mass for at 14790.5 Da in the protein expression profiles.

Using this approach, six significant protein markers from the normal crypt and adenomas comparisons were identified from the cellular homogenates. These identified protein markers are serine protease inhibitor kazal type 3 (Spink3), calmodulin, cysteine-rich intestinal protein (CRIP), calgranulin A, calcyclin, and galectin-2. Spink3, CRIP, and calmodulin were overexpressed in the normal crypt compared to the adenomas, and were more abundant in the top of crypts vs. the bottom of crypts. Calgranulin A, calcyclin, and galectin-2 were found overexpressed in the adenomas. The average molecular weight, the sequence coverage from the MS/MS analyses, the overexpressed region and brief description of the biological functions of these protein markers are shown in Table 5.

Immunohistochemistry of Identified Protein Markers

Immunohistochemistry was performed to confirm the differential expression of the identified protein markers. From the mass spectrometric analysis, CRIP was decreased in colorectal adenomas vs. normal crypts. Within the crypts, CRIP was expressed relatively higher in the top compartment compared to the bottom compartment. These results were consistent with

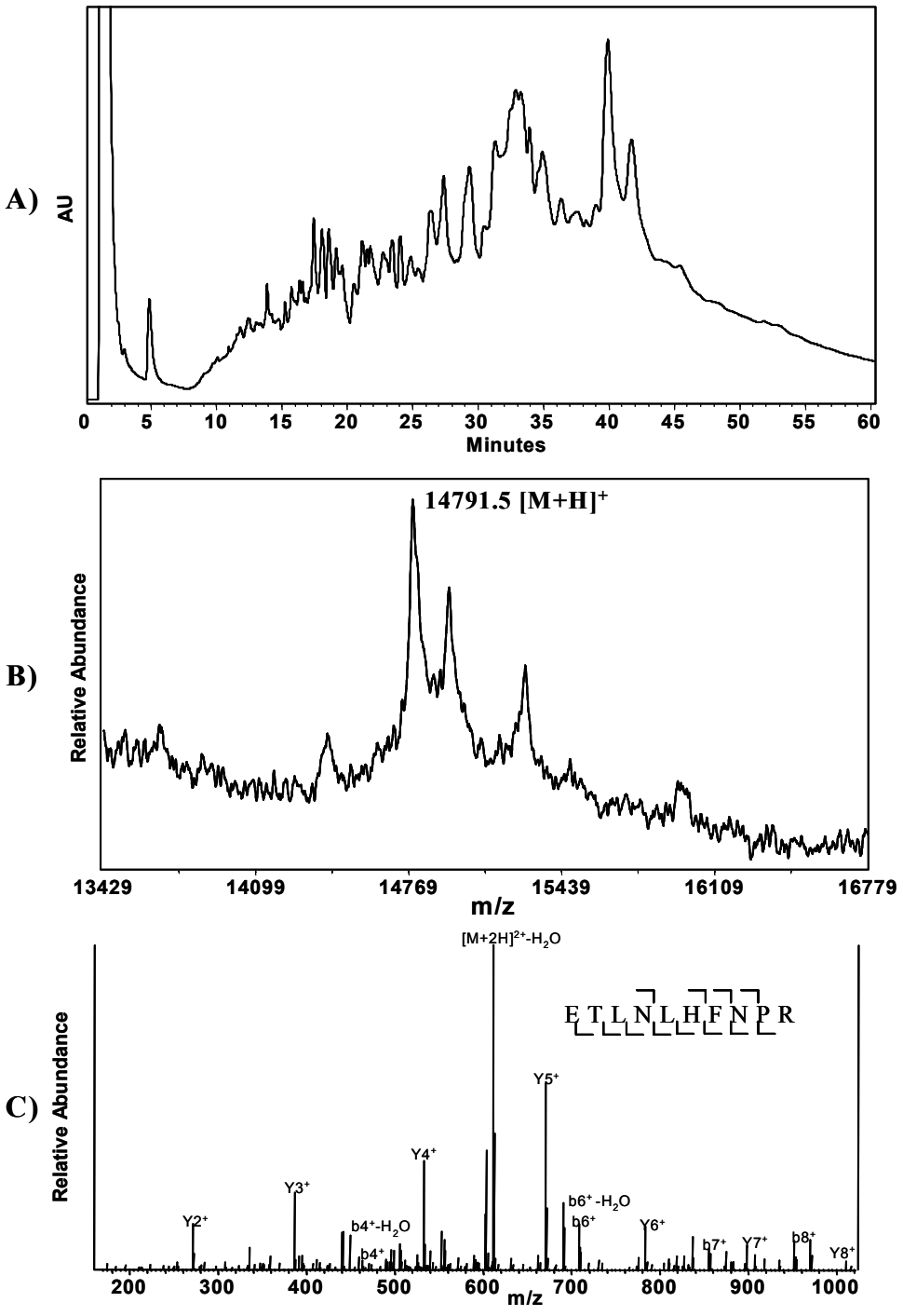


Figure 21: The identification of galectin-2. A). UV trace (214 nm) for the HPLC of the polyp homogenate; B) MALDI mass spectra of the fraction which contains the target protein; C) MS/MS of the tryptic peptide at m/z 621.1 (z = 2+) of galectin-2. A total of four peptides were sequenced for galectin-2 in the trypsin digestion of the HPLC fraction.

Table 5: The list of identified protein markers resulted from the comparison between the crypt base and adenomas.

Protein Name	Molecular Weight (Da)	MS/MS sequence coverage (number of sequenced peptides)	Higher Expression	Function
Spink3	6120	60% (2)	Bottom Crypt	Exhibits anti-trypsin activity
CRIP	8419	60% (4)	Bottom Crypt	Intracellular zinc transport
Calmodulin	16792	36% (3)	Bottom Crypt	Regulation of various enzyme activity by Ca^{2+}
Calgranulin A	10165	74% (3)	Adenoma	Regulates immune reactions
Calcyclin	9961	34% (3)	Adenoma	Interacts with annexin II in a calcium-dependent manner
Galectin-2	14791	39% (4)	Adenoma	Binds β-galactoside

the immunohistochemical staining for CRIP (Figure 22 A and B). Calcyclin was found to be increased in the adenomas vs. normal crypts from direct protein profiling and thus was confirmed by the immunostaining results (Figure 22 C and D). Similarly, calgranulin was increased in the adenomas vs. normal crypts and the mass spectrometry analysis was consistent with the immunostaining results (Figure 22 E and F).

Discussion

In this study, epithelial cells from the top third and bottom third compartments of normal mouse colonic crypt and adenomatous cells from *Apc*$^{Min/+}$ mice polyps were obtained using laser capture microdissection. The protein profiles were directly obtained from these microdissected cells using MALDI-TOF MS. These profiles were analyzed using WFCCM program and statistically significant protein markers for each group were obtained. Using hierarchical clustering analysis, the protein profiles from the three groups were accurately classified from each other. A set of highly statistically significant protein markers were identified and their differential expressions were confirmed with immunohistochemistry.

Dividing the crypts into top and bottom compartments for proteomic comparisons with adenomas is a critical first step in elucidating protein changes specific to the initial stages of colonic neoplasia. The bottom crypt region is widely believed to be the site where colon cancer cells initiate due to the stem cells locating in this compartment [54, 55]. The daughter cells rapidly proliferate in the bottom crypt and migrate up along the crypt, gradually losing their capacity to divide. When these cells reach to the top crypt, they undergo apoptosis and shed into the lumen. Any genetic alterations are lost in daughter cells as they consistently move out of the crypt. Only the stem cells are in a steady state of cell growth and permanently tethered to a niche at the bottom crypt. Oncogenic mutations accumulate in the bottom compartment which is considered as the target cells during colonic neoplasia transformation. Therefore, providing the

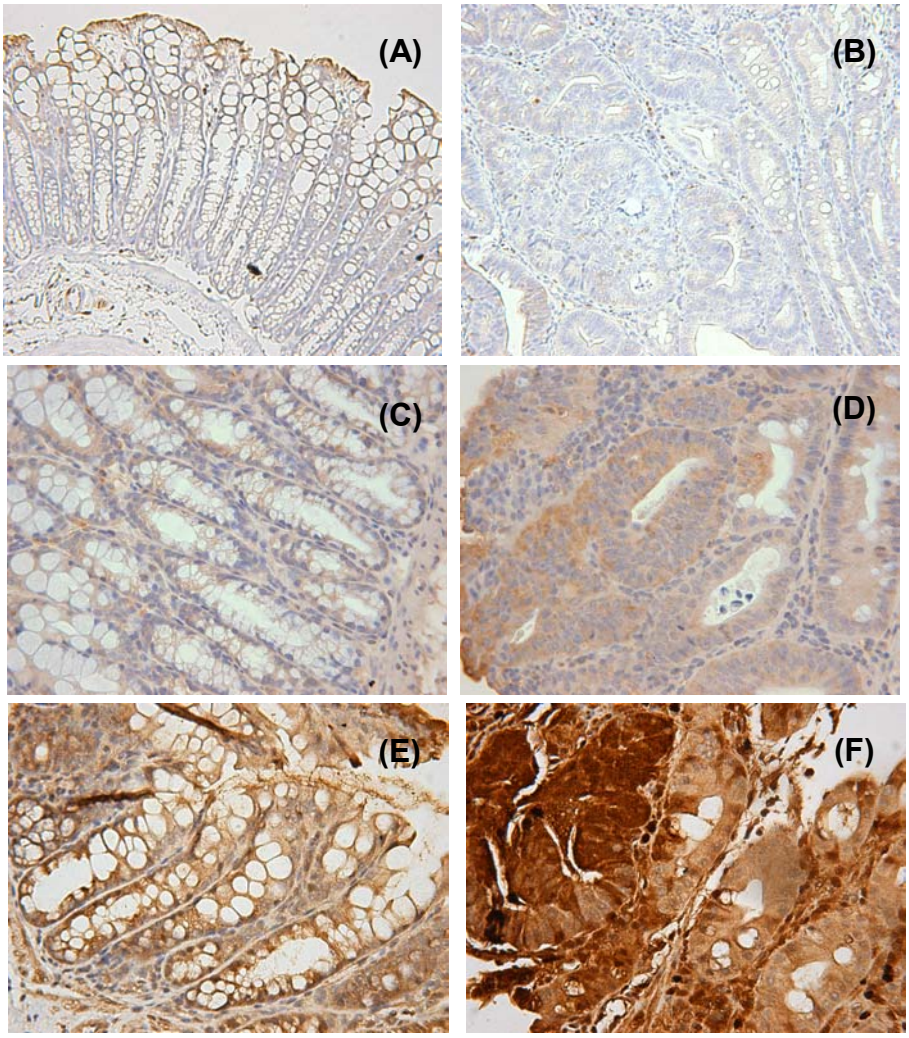


Figure 22: The immunohistochemistry results are consistent with the proteomic findings. CRIP was expressed higher in the top of the crypt (A) but not in colonic adenomas (B). The presence of calcyclin was lower in normal crypts (C) compared to adenomas (D). Calgranulin A expression was also lower in normal crypts (E) compared to colonic adenomas (D). The panel (A) and (B), (C) and (D), (E) and (F) were from the in same tissue section. The magnification power for (A) and (B) were 20X and (C)-(F) were 40X.

proteomic pattern specific to the bottom crypt compartment has a significant value in serving as an accurate reference for the molecular transition in CRC tumorigenesis.

In this study, the different protein expression profiles from the top and bottom of normal crypts were demonstrated using the WFCCM method and hierarchical clustering method. The near perfect classification of the 130 top and bottom crypt protein profiles were achieved using both of the statistical methods. This represents the first time that top and bottom crypts can be accurately classified from each other based on their proteomic pattern differences.

Perfect classification was achieved for both cases for the proteomic pattern comparisons of adenomas vs. top and bottom of normal crypts. We chose to focus on the comparison between the crypt base and adenomas because this comparison is more relevant in elucidating the protein markers associated with colorectal tumorigenesis. Both cells at the crypt base and neoplasia cells constantly differentiate and proliferate, with the difference being the former is well controlled and the latter escapes apoptosis regulation. Thus, the protein markers derived from this comparison, as shown in Figure 19B, hold promise in revealing protein markers specific to highly proliferative cells which also are malignant.

Some of the statistically significant protein markers were identified through sequence analyses using tandem mass spectrometry. Their differential expressions were further confirmed using immunohistochemistry. Protein markers overexpressed in colonic neoplasia are also in agreement with literature. For example, calcyclin, found up-regulated in the $Apc^{Min+/-}$ adenomas in this study, has also been reported as overexpressed in human colonic neoplasia [66, 67]. Calcyclin belongs to the S100 protein family and contains two EF-hand calcium-binding motifs. It has been implicated as being involved in cell regeneration [68], differentiation [69], and metastasis [70, 71]. Calcyclin also has been suggested to play a role in intracellular calcium homeostasis, signaling and ion transport [72]. The overexpression of calcyclin has been reported

in human melanoma [70], squamous cell carcinoma of the oral mucosa [73] and human breast cancer [74].

Another S100 protein found upregulated in the *Apc*$^{Min+/-}$ adenomas is calgranulin A, whose overexpression has also been reported in previous proteomic study in azoxymethane-treated mouse colorectal tumor [75] and human colorectal cancer [76]. Our immunostaining confirms its elevated expression in the *Apc*$^{Min+/-}$ colonic adenomas. Calgranulin A binds with calgranulin B and forms a complex called calprotectin [77]. As a calcium and zinc binding protein, calprotectin is a multipotent molecule and can inhibit several zinc-dependent metalloproteinase by sequestration of zinc [78]. Calprotectin is found abundantly in neutrophils and monocytes. Calprotectin's level in plasma, stools or cerebrospinal fluid is increased when leukocytes are activated [79]. Patients with gastrointestinal cancers have been found with higher fecal calprotectin levels than control patients [80]. Though calprotectin is not a marker specific to colon cancer, it is has been routinely used as a marker for patients with inflammatory bowel disease in Norway and several other centers in Europe [81].

Galectin-2 was also found to be overexpressed in the *Apc*$^{Min+/-}$ adenomas compared to normal crypts. Galectin-2 belongs to the galectin family which binds specifically to β-galactos-containing oligosaccharides [82]. Galactin-1 and galectin-3 have been shown to be upregulated in human colon cancer [83], pancreatic tumor [84], and head and neck squamous cell carcinoma [85] using immunohistochemistry and western blotting. Increased galectin-3 expression correlates with decreased survival in colorectal cancer[86, 87]. Therefore, though the functions of galectin-2 in CRC have not been studied extensively, it shows a promising role in colorectal carcinogenesis.

As a valuable discovery tool, LCM combined with MALDI MS technology eliminates the need to know in advance the specific proteins that are changed in the comparative analyses. Some protein markers which have not been previously reported as associated with CRC were also found in this study, including CRIP, SPINK3, and calmodulin. The biological functions of these

protein markers need to be further explored for their potential associations with CRC development.

Some limitations and future developments need to be addressed. First, the identified proteomic patterns should be further validated in human colonic polyps. Though *Apc*$^{Min+/-}$ mice polyps have similar genotype and phenotype with human familial adenomatous polyposis (FAP), other types of colonic cancer also exist for human colonic malignancies, such as hereditary non-polyposis colon cancer (HNPCC) and sporadic colon cancer. Whether the protein makers found in this study are the same for different types of CRC should be validated. Second, although the direct tissue protein profiling approach provides highly accurate and specific protein markers for neoplastic cells, it is an invasive technique and limited for patients who have colonic biopsies, thus making it difficult to use as a screening tool. However, these neoplastic protein markers detected in tissue could degrade into serum or feces. Detecting the specific fragments or intact protein markers from the serum or fecal matter would be more applicable for colorectal cancer screening.

In conclusion, our data show that the proteomic difference present within minute tissue structure like normal murine crypts can be obtained using LCM and MALDI MS. Colonic neoplastic protein markers can be elucidated. Further functional analysis of the biological activities of the protein markers will advance our understanding of the colorectal tumorigenesis and will assist CRC diagnosis and treatment.

Materials and Methods

Chemicals

Sinapinic acid, ammonium hydrogen carbonate, dithiothreitol, iodoacetamide and phenylmethanesulfonyl fluoride (PMSF), were purchased from Sigma (St. Louis, MO, USA). Sucrose and Tris were purchased from J. T. Baker (Phillipsburg, NJ, USA). Formic acid (98-

100%) was from EM Science (Merck, Darmstadt, Germany). Sequencing-grade trifluoroacetic acid (TFA) was purchased from Burdick and Jackson (Muskegon, MI, USA) and HPLC grade acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Bovine sequence-grade trypsin was purchased from Promega (Madison, WI, USA).

Laser Capture Microdissection

A total of four wild-type and eight *Apc*$^{Min+/-}$ adult mice with C57/Bl6 background were used for the protein expression profiles analyses. The *Apc*$^{Min+/-}$ polyps were a gift from Dr. Joanna Groden. The frozen colon and polyps were partially embedded in OCT and five micron frozen sections were cut at –15^{o}C using a cryostat. Each tissue section was transferred onto a glass slide and immediately dehydrated as followed: 70% ethanol 1 minute, 95% ethanol 1 min, 100% ethanol 1 min (2X), xylene 2.5 min (2X) and air dried for 5 min. To investigate the number of cells captured per laser shot, a Hematoxylin and Eosin stained section was prepared. The H&E sections showed the polyps were advanced adenomas with areas of cribriform architecture indicative of high grade dysplasia. The LCM was performed using an Arcturus PixCell II LCM system and CapSure caps (Arcturus, Mountain view, CA). The diameter of the LCM laser beam was focused at 30 micrometer. The laser was carefully aimed and fired at either the bottom or top third of the crypts and the specific cells of interested were microdissected onto the LCM film. The adenoma cells from the *Apc*$^{Min+/-}$ mice were microdissected in the similar manner. Approximately 100 cells were microdissected per mass spectrometry analysis.

MALDI MS Sample Preparation and Analysis

The LCM film with attached cells was gently peeled off from the LCM cap and adhered to a MALDI target plate with double-sided tape (Digi-Key, Thief River Falls, MN). Sinapinic acid at 20 mg/ml in 5/5/0.3 v/v/v acetonitrile/water/trifluoroacetic acid was used as the matrix

solution. Under microscope visualization, the matrix solution was carefully deposited only on the captured cells with a pulled glass capillary. The MS analysis was performed in the linear mode under optimized delayed ion extraction conditions using a DE-STR MALDI TOF mass spectrometer (Applied Biosystems, Framingham, MA). Three hundred laser shots were used to generate one mass spectrum from the captured cells. Constant instrumental settings including laser energy were used for all the analysis.

Statistical Analysis

The mass spectra were baseline corrected, normalized and the peaks were aligned across all the acquired mass spectra using custom algorithms developed at Vanderbilt University. Statistically significant MS signals were selected based on the Kruskal-Wallis test, Fisher's exact test (dichotomizing the expression level as present or not), the permutation t-test, Significance Analysis of Microarrays (SAM), Weighted Gene Analysis (WGA), and the modified info score method. The cutoff points for each method were $p<0.001$, $p<0.001$, $p<0.001$, 2.75, 2 and 0, respectively. A protein that met at least three of these six selection criteria was selected on the final list. The Weighted Flexible Compound Covariate Method (WFCCM) [64] was employed in the class-prediction model based on the selected proteins to verify whether the proteomic patterns could be used to classify tissue samples into adenoma, crypt tops and crypt bottoms. The WFCCM method reduced the dimensionality of the problem by using a new covariate obtained as a weighted sum of the most important predictors and combining the most significant proteins associated with the biological status from each analysis method. We estimated the misclassification rate using the leave-one-out cross-validation class prediction method based on the WFCCM. The agglomerative hierarchical clustering algorithm was also applied to investigate the pattern among the statistically significant discriminator proteins as well as the biological status using Eisen's software.

Protein Identification

Three mice were sacrificed by CO_2 inhalation and the crypts were isolated from the colons as the technique described previously [88]. The isolated colonic crypt pellet (100 mg) was suspended in 1ml of homogenization buffer (0.25M sucrose, 0.01M Tris-HCl and 0.1mM phenylmethylsulfonyl fluoride) and homogenized with a Duall homogenizer. The lysate went through a series of centrifugation steps (10 min at 680 g, 10 min at 10,000g and 1h at 100,000g). Then 250 μl of 0.3 M NaCl/0.01 M Tris-HCl solution was used to extract protein of the pellet from the 680g centrifugation pellet with multiple bursts of a Kontes motor-driven polyproylene pestle for one minute and the solution was centrifuged at 16,000g for ten minutes. All the procedures were done at 4°C.

The supernatant from the 100,000g centrifugation and salt extraction of the pellet was further separated with a polymeric reverse phase HPLC column (259VHP5415, Vydac, Hesperia, CA). The solvent A was 0.1% trifluoroacetic acid and solvent B was 0.085% trifluoroacetic acid in acetonitrile. The flow rate was 1 ml/min and the gradient was 5 min at 5% B, then in 75 min to 60% B, then in 10 min to 95% B. The fractions were collected every minute up to 80 minutes. The HPLC fractions containing the proteins which had the target molecular weight were found using MALDI mass spectrometry. For trypsin digestion, the HPLC fractions of interest were lyophilized and reconstituted with 10 μl of 0.4 M ammonium hydrogen carbonate and 5 μl of 45 mM dithiothreitol solution, followed by incubation at 60°C for 15 min. The solution was set in the dark for 15 min after adding 5 μl of 100mM iodoacetamide. 1 μl of 1 μg/μl modified porcine trypsin (Promega, Madison, WI) was added and incubated at 37^{o}C for 4h. The tryptic peptides were sequenced with nanoelectrospray MS/MS using an MDS SCIEX Qstar QTOF mass spectrometer (Applied Biosystems, Toronto, Canada). For the fractions containing complex protein mixtures, the tryptic peptides were sequenced using capillary LC-MS/MS with LCQ Deca

(ThermoFinnigan, San Jose, CA) and the sequenced peptides were searched using the SEQUEST algorithm.

For the identification of adenoma markers, 300 mg of the polyp was homogenized and centrifuged according to the procedure described above. The nuclear fraction was extracted with 450 µl of 0.3M NaCl solution. 400 µl of the salt extract solution was separated over a C4 Vydac column (214MS5215) at 40 °C using the Waters HPLC system. Using the same mobile phase described above, the gradient started from 5% B for 5 min, to 25% B in 15 min, then to 70%B in 55 min, and finally to 95% B in 10 min. The fractions were analyzed using the protocol described above.

Immunohistochemistry

Five micron sections of paraffin-embedded tissues were placed on charged slides and the paraffin removed. The sections were rehydrated and placed in heated Target Retrieval Solution from DakoCytomation (Carpinteria, CA) for 20 minutes. For calgranulin-A immunostaining, endogenous peroxidase was neutralized with 0.03% hydrogen peroxide and samples were treated with diluted rabbit serum prior to primary antibody addition. Tissues were incubated with goat anti-Calgranulin A (m-19) from Santa Cruz Biotechnology (Santa Cruz, CA) diluted 1:500 for 30 minutes. Sections without primary antibody served as negative controls. The Vectastain ABC Elite (Vector Laboratories, Inc) System and DAB+ (DakoCytomation) was used to produce localized, visible staining. Slides were lightly counterstained with Mayer's hematoxylin, dehydrated and coverslipped. For the CRIP immunostaining, after the sections were rehydrated and placed in heated Target Retrieval Solution (DakoCytomation), endogenous peroxidase activity was inactivated with 0.03% hydrogen peroxide followed by a casein-based protein block (DakoCytomation) for nonspecific staining blocking. The sections were incubated with rabbit anti-CRIP diluted 1:50 for 120 minutes. Sections without primary antibody served as negative

controls. The Dako Envision+ System, rabbit/HRP (DakoCytomation) was applied for 30 minutes. DAB+ was used to produce specific, visible staining. The slides were lightly counterstained with Mayer's hematoxylin, dehydrated and coverslipped. The immunohistochemistry was performed by Kelly Parman at the VUMC Immunohistochemistry Core lab.

CHAPTER IV

PROTEOMIC PATTERNS AND PREDICTION OF GLOMERULOSCLOROSIS AND ITS MECHANISMS

Introduction

As key structures of kidney nephrons, glomeruli are tiny round structures in the kidney cortex composed of capillary blood vessels that are actively involved in the filtration of the blood to form urine. The scarring of the glomeruli is referred as glomerulosclerosis, which is a common lesion for a variety of renal diseases, such as idiopathic focal segmental glomerulosclerosis (FSGS). The incidence of FSGS is increasing and the recurrence rate is high for post renal transplantation, making it one of the primary glomerular diseases. Progression of FSGS is postulated to develop from early diffuse podocyte injury preceding overt sclerosis [89, 90], likely involving multiple complex mechanisms. With the advancement of proteomic techniques [91, 92], simultaneous examination of hundreds of proteins related to kidney disease holds promise in unraveling novel underlying mechanisms of progression and thus identifying possible targets for intervention in FSGS.

The focal segmental nature of sclerosis raises the question of whether the remaining nonsclerotic glomeruli at a given time point in FSGS already are programmed to sclerotic pathways, or alternatively have less prosclerotic activation and thus may be more susceptible to therapy. We therefore sought to examine proteomic profiles from the nonsclerotic glomeruli in the FSGS kidney, and compare to normal baseline and glomeruli with established sclerosis. The following specific aims are developed for this application:

Specific Aim

To obtain proteomic patterns to begin classifying glomerulosclerosis at the protein level and to advance the understanding of glomerulosclerosis mechanisms.

Results

Protein Profiles of Normal, Non-sclerotic and Sclerotic Glomeruli

LCM allowed precise dissection of only glomeruli without surrounding tissue under LCM phase microscopy from surrounding cortex under LCM microscope without staining, as shown in (Figure 23). Each MALDI mass spectrum was obtained from an average of five microdissected glomeruli (~200 cells) averaging signals from 250 consecutive laser shots. From the six rats, a total of 60, 30, and 60 mass spectra were obtained from the normal, nonsclerotic and sclerotic glomeruli, respectively. Approximately two hundred protein signals were detected per spectrum in the mass range of 2,000-70,000 Da, with the signals under 20,000 Da yielding the best resolution. Differentially expressed signals were found among the three different classes of glomeruli, as shown in Figure 24.

Glomerular Proteomic Pattern Comparisons

Total of 1473 distinct peaks across all the spectra were obtained. Using the weighted flexible compound covariate method (WFCCM) statistical analysis, we were able to classify the normal and sclerotic glomeruli proteomic pattern with 99.2 % accuracy using the top 102 differentially expressed MS signals. Similarly, 96.7% classification accuracy was obtained for the comparison of normal glomeruli vs. nonsclerotic glomeruli using the top 166 differentially expressed MS signals. A classification accuracy of 97.8% was obtained for nonsclerotic glomeruli from sclerotic glomeruli using the top 84 differentially expressed MS signals (Table 6).

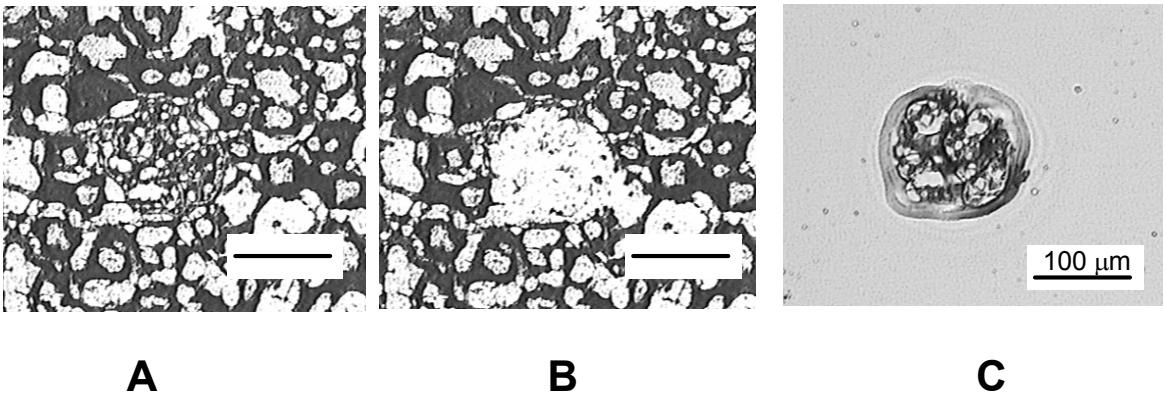


Figure 23. The process of laser capture microdissection of a normal glomerulus. A). The normal glomerulus can be clearly recognized. B). The remaining tissue section after the glomerulus was microdissected. C). The microdissected glomerulus on the LCM film. Each microdissected glomeruli section contains approximately 50 cells.

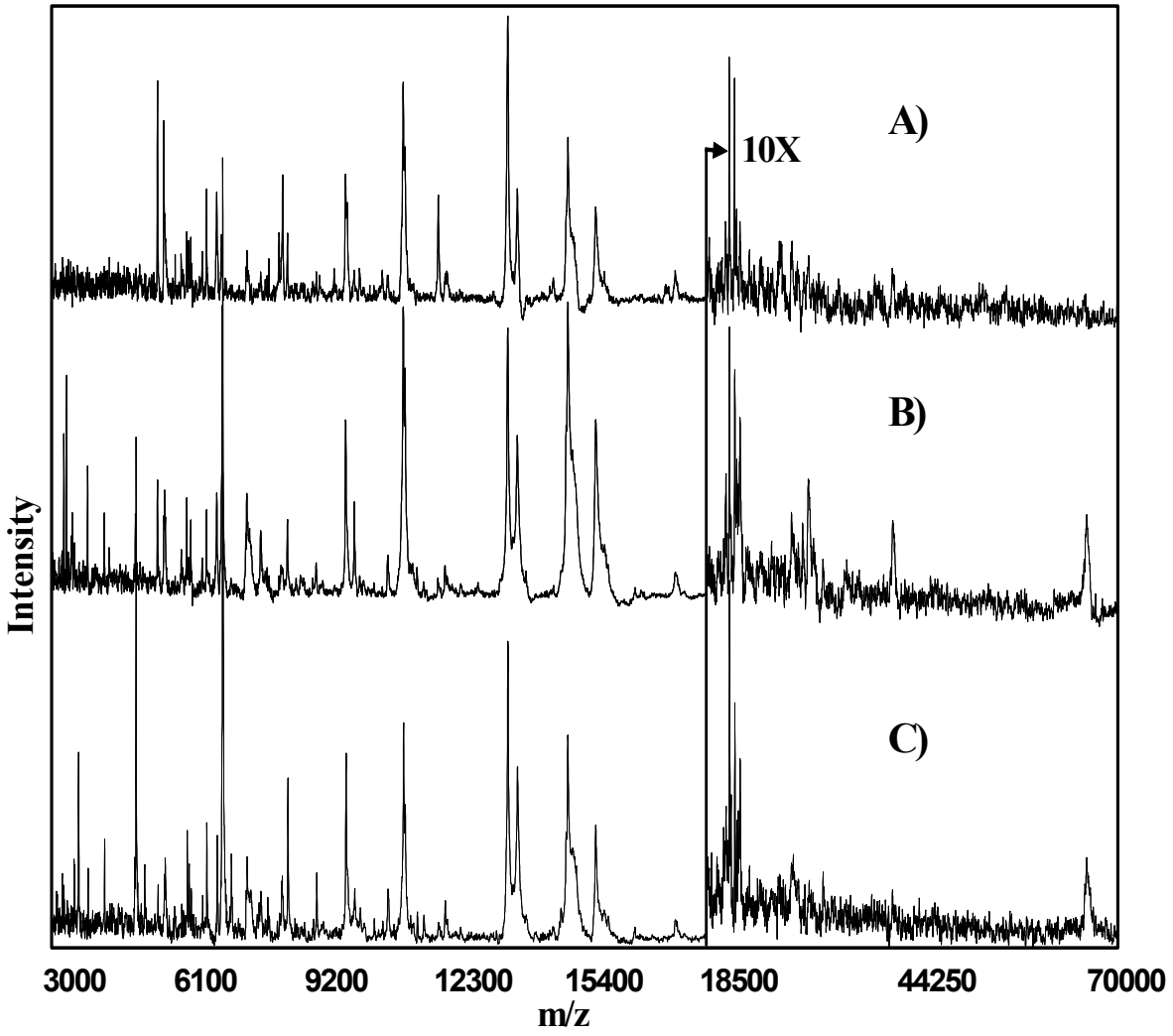


Figure 24. Mass spectra obtained from three groups of glomeruli: A) normal baseline glomeruli, B) nonsclerotic glomeruli, and C) sclerotic glomeruli.

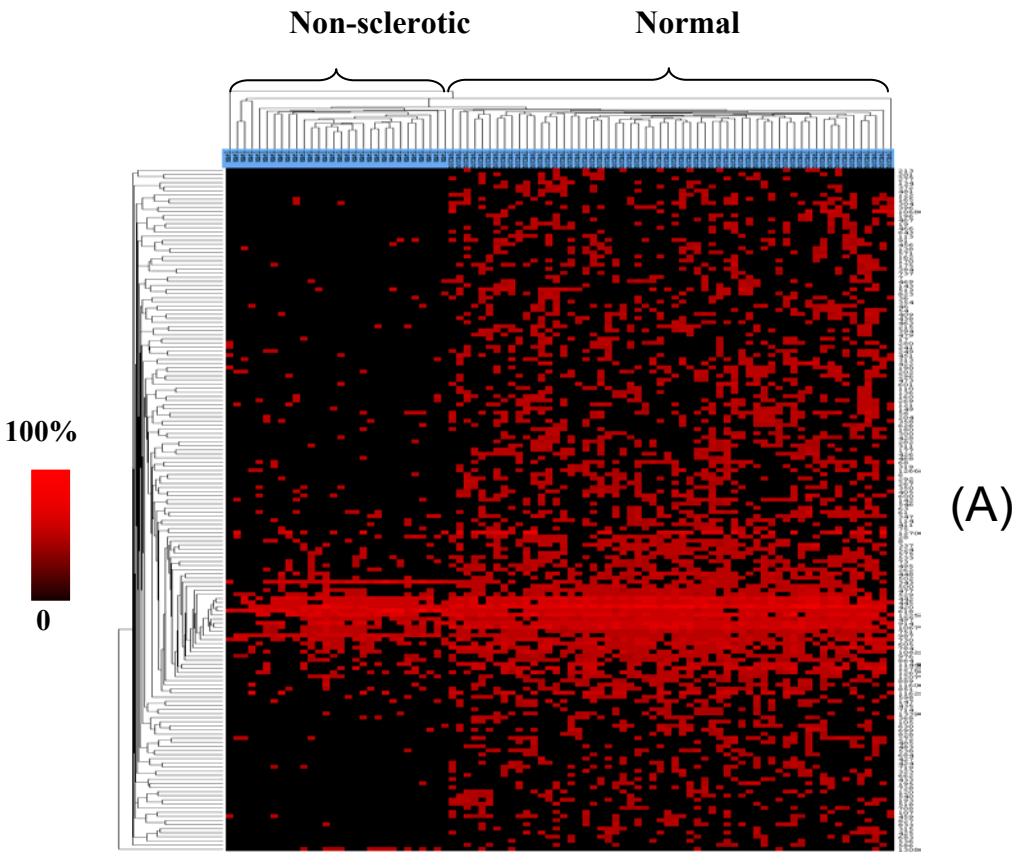
Table 6. The weighted flexible compound covariant method (WFCCM) was used to select the statistically significant peaks that allowed classification of the three groups of samples according to their proteomic patterns. The misclassification rate was calculated using the leave-one-out cross-validation class prediction method[64].

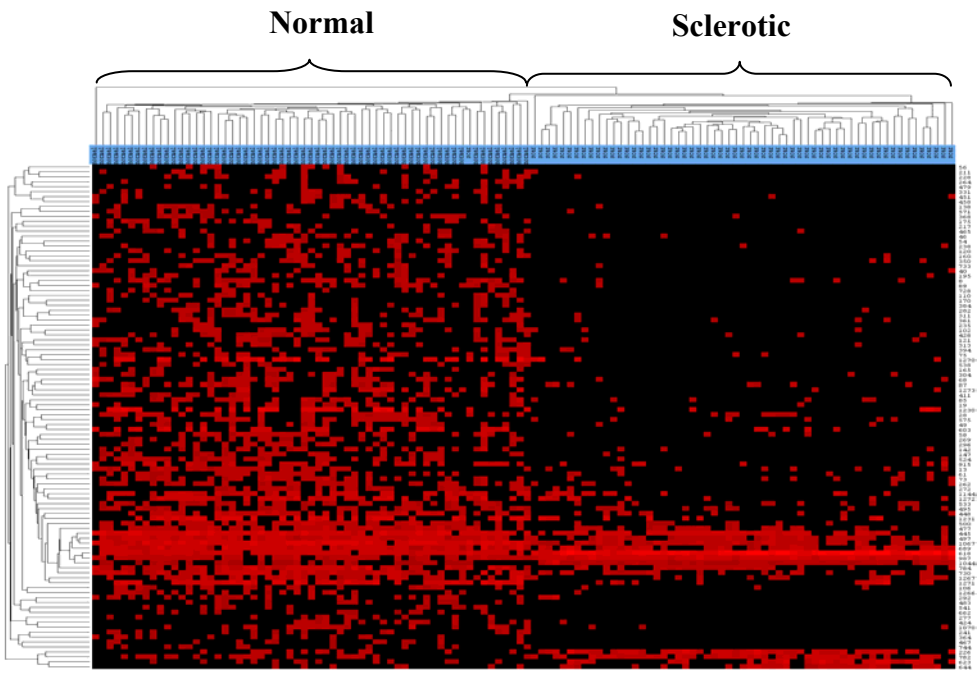
Classification (sample size)	No. of differentially expressed peaks	No. of misclassified samples (%)	Probability of random permutations with misclassifications
Normal Glomeruli (60) vs. Sclerotic Glomeruli (60)	102	1 (99.2)	< 0·0001
Non-sclerotic Glomeruli (30) vs. Normal Glomeruli (60)	166	3 (96.7)	< 0·0001
Non-sclerotic Glomeruli (30) vs. Sclerotic Glomeruli (60)	84	2 (97.8)	< 0·0001

The agglomerative hierarchical clustering algorithm was used to investigate the protein expression patterns among the significant differentially expressed proteins with Eisen's software. The selected proteomic pattern distinguished all the normal glomeruli from nonsclerotic glomeruli with 100% classification accuracy. A classification accuracy of 99.2% was obtained in distinguishing normal vs. sclerotic glomeruli and 92.2% classification accuracy for nonsclerotic glomeruli vs. sclerotic glomeruli (Figure 25). The overall proteomic pattern of nonsclerotic glomeruli was more similar to sclerotic than to normal glomeruli ($P<0.0001$).

Identification of Thymosin β4 and Its Expression In Vivo and In Vitro

The protein markers that statistically most significantly contributed to differential classification of our three classes of glomeruli were targeted for identification. After the tissue homogenization and HPLC separation steps, three fractions containing the peak of m/z value 4963.76, one of the target proteins, were found using MALDI MS (Figure 26A). With LC-MS/MS analysis of the resulting tryptic peptides, thymosin β4 was identified as the target protein marker. Multiple MS/MS spectra were found to be consistent with the thymosin β4 tryptic peptides: $[TETQEKNPLPSK]_{20-31}$, $[KTETQEKNPLPSK]_{19-31}$, $[TETQEKNPLPSKETIEQEK]_{20-38}$, and $[KTETQEKNPLPSKETIEQEK]_{19-38}$. The sequest cross correlation scores for these sequences were 3.42, 3.89, 4.95 and 4.69, respectively, showing a strong correlation between the MS/MS spectra and the amino acid sequences. As an example, the MS/MS spectra for $[TETQEKNPLPSK]_{20-31}$ is shown in figure 26B. These sequences composed 44% of the total amino acid sequence of thymosin β4. Considering the previous reported N-terminal acetylation [33], the theoretical average molecular weight for thymosin β4 is 4963.5 Da, which matches well with the target protein molecular weight. The signal intensities of thymosin β4 from the mass spectra for normal glomeruli, nonsclerotic glomeruli and sclerotic glomeruli are shown in Figure


Non-sclerotic
Normal
100%
0
(A)



(B)

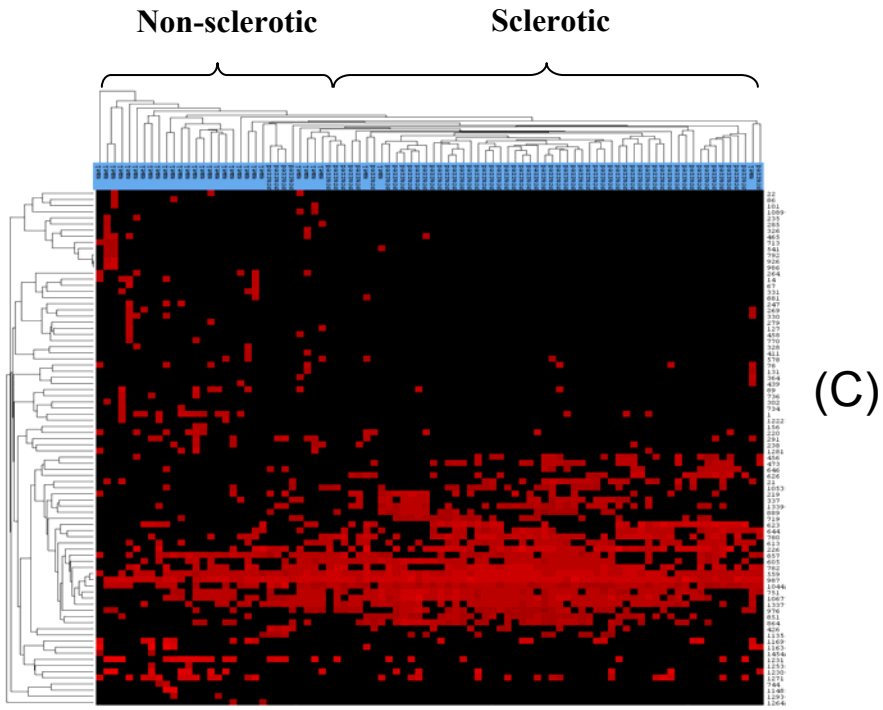


Figure 25. Hierarchical cluster analysis of three different groups of glomeruli based on the top differentially expressed protein patterns. A) Total of 30 non-sclerotic glomeruli and 60 normal glomeruli protein profiles were accurately classified; B) One out of 60 normal glomeruli and 60 sclerotic glomeruli protein profiles was misclassified; C) Seven out of 30 non-sclerotic glomeruli and 60 sclerotic glomeruli protein profiles were misclassified.

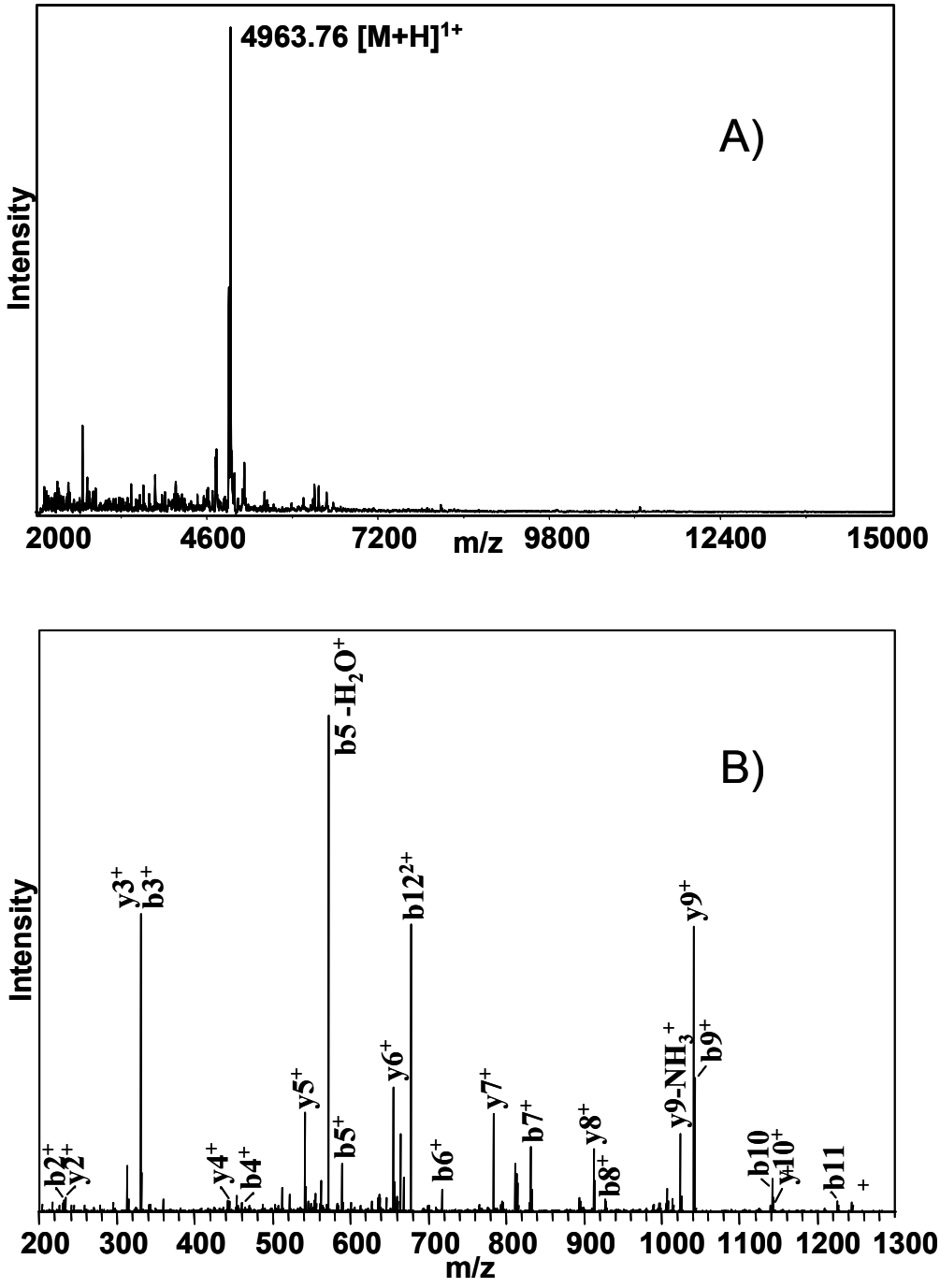


Figure 26. Identification of thymosin β4 as a marker for sclerosis. A) The MALDI mass spectrum of a HPLC fraction that contains the target peak at 4963.76 m/z. B) ESI MS/MS spectrum consistent with one of thymosin β4's tryptic peptides, [TETQEKNPLPSK]20-31.

27. Thymosin β4 expression levels were increased approximately 3-fold from normal to sclerotic glomeruli based on the MS intensities.

Immunohistochemistry was performed to confirm the increased level of thymosin β4 in glomerulosclerosis. Thymosin β4 was increased in sclerotic glomeruli vs. non-sclerotic or normal glomeruli (Figure 28). Furthermore, thymosin β4 was found predominantly expressed in endothelial cells identified by serial section staining with RECA-1 (Abcam Inc, Cambridge, MA), while mesangial cells, stained with anti-rat Thy-1 (BD Pharmingen, San diego, CA) were negative for thymosin β4. Podocytes, identified anatomically, and macrophages, double stained with ED1 (Biosource) were also negative for thymosin β4.

Further analyses were performed *in vitro* assessing thymosin β4 expression in two different glomerular cell lines (Figure 29). Western blots from cultured glomerular endothelial (GEN) and podocyte cells confirmed endothelial expression of thymosin β4, with no protein detected in podocytes. GEN showed a strong immunoreactive band at 4.9 kDa corresponding to the expected molecular weight of thymosin β4. Mouse spleen and muscle, used as positive and negative controls of thymosin β4 immunohistochemistry and Western blot expression, showed expected results.

Thymosin β4 Effect on Sclerosis Mechanisms

Whether modulation of thymosin β4 mRNA affected responses to Ang II were assessed. Angiotensin stimulates PAI-1 *in vivo* and *in vitro* [93, 94]. We therefore assessed PAI-1 expression in response to Ang II in GEN. To investigate the functional role of thymosin β4 in sclerosis, we designed siRNA and control siRNA for thymosin β4 and transfected them into GEN. Thymosin β4 protein expression was successfully knocked down by approximately 90% using siRNA, with equivalent efficacy of all four siRNAs tested. Scrambled control RNA had no effect. We next assessed effect of downregulated thymosin β4 on the Ang II–induced prosclerotic

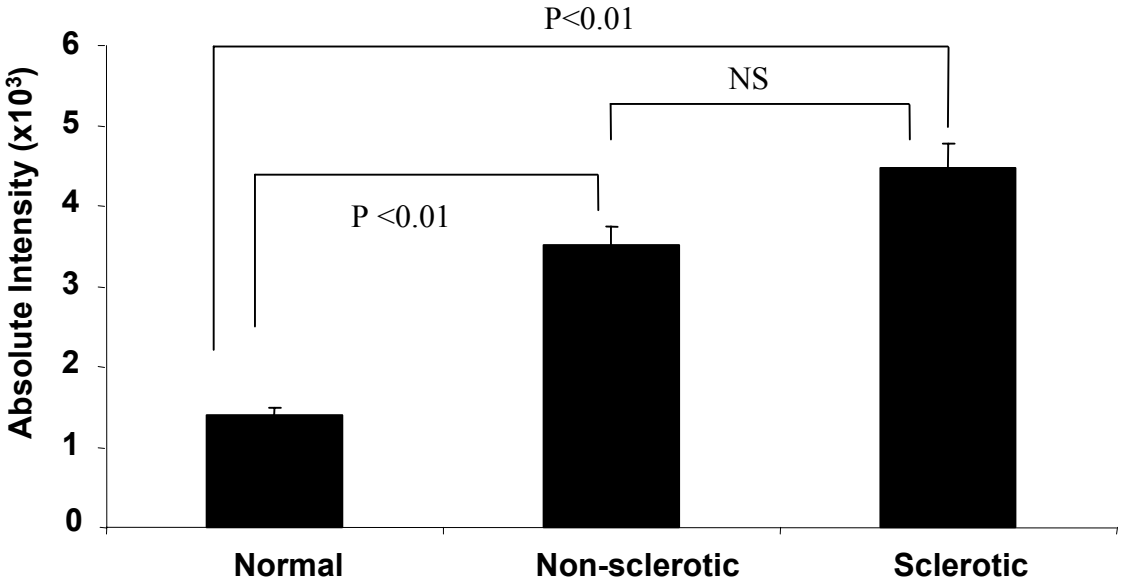


Figure 27: Thymosin β4 intensity levels based on MS signal intensities in the three different glomerular groups. Statistical significance of the differences was determined by using a two-tailed Student's *t* test for normal vs. nonsclerotic glomeruli (p <0.01), normal vs. sclerotic glomeruli (p < 0.01), and nonsclerotic vs. sclerotic glomeruli (p > 0.05) comparisons. The error bars correspond to 95% confidence intervals.

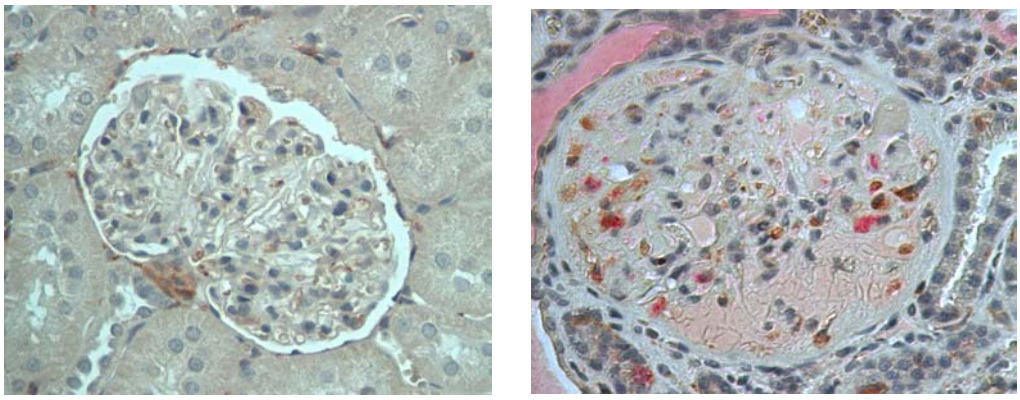


Figure 28. Thymosin β4 (brown) was not detectable in normal glomeruli (left), but was increased in sclerotic glomeruli (right). Staining was not present in macrophages (red), staining on serial sections with endothelial and mesangial markers (not shown) confirmed predominantly endothelial localization (anti-thymosin β4, brown; anti-CD68, red).

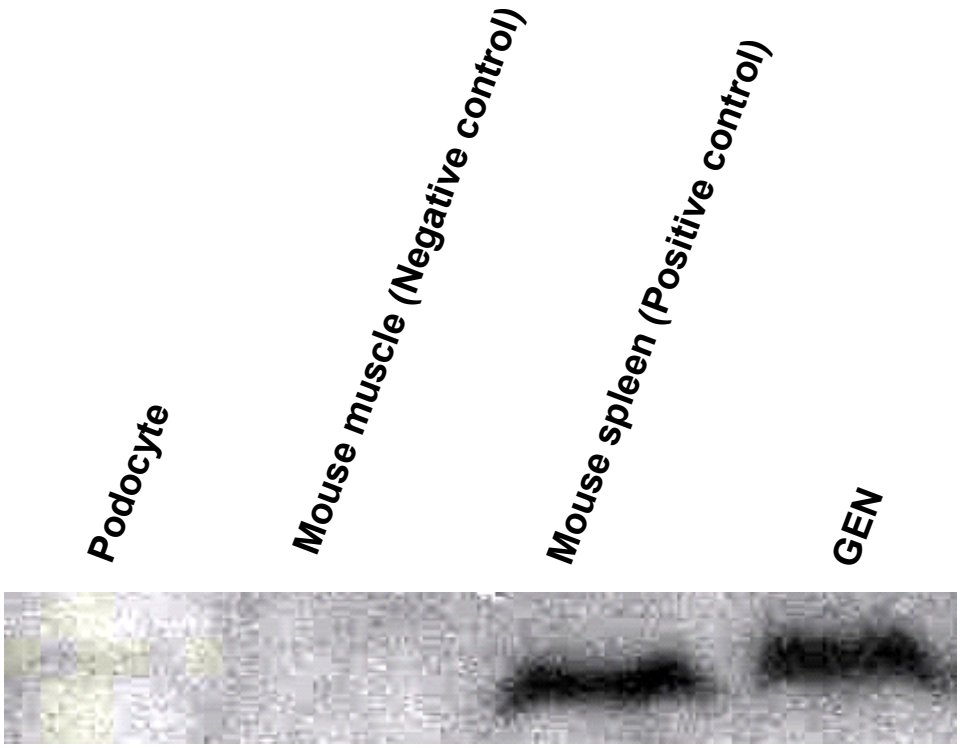


Figure 29: Western blot of thymosin β4 expression in cultured glomerular endothelial cells and podocytes. Mouse muscle and spleen tissue were used as negative and positive controls, respectively.

response, by assessing expression of PAI-1 in these cells. Neither siRNA or control siRNA affected baseline PAI-1 expression. Angiotensin II (10^{-6} M) increased thymosin β4 over baseline, and concurrently dramatically upregulated PAI-1 in normal GEN, while transfection with siRNA for thymosin β4 significantly decreased the angiotensin II - induced PAI-1 expression. The control siRNA had no effect on the Ang II – induced PAI-1 expression (figure 30).

Discussion

Our results show that LCM and MALDI-TOF MS can accurately classify the proteomic profiles of normal vs. nonsclerotic and vs. sclerotic glomeruli. Using overall proteomic pattern similarity comparisons, nonsclerotic glomeruli were found to have more similarities to the sclerotic glomeruli than to the normal glomeruli. Thymosin β4 was identified as one of the key differentially expressed protein markers upregulated both in nonsclerotic and sclerotic glomeruli in the 5/6 nephrectomy model. Immunohistochemistry confirmed thymosin β4 elevated expression levels in sclerotic glomeruli, localized particularly to endothelial cells. Using RNAi technology, thymosin β4 was found to regulate a key pathway in sclerosis, namely angiotensin II-induced PAI-1 expression.

Normal and sclerotic glomeruli represent the two extremes of evolution of glomerulosclerosis. In this study, we confirmed, as expected, that proteomic patterns for these normal and sclerotic glomeruli are markedly different. Of greater interest is our finding that nonsclerotic glomeruli in the setting of progressive sclerosis also have an altered proteomic profile, more closely related to sclerotic than normal glomeruli. Foot process effacement occurs in all glomeruli in FSGS, regardless of whether sclerosis is present in the glomerulus or not. Thus, podocyte-related molecules would be expected to be altered in both nonsclerotic and sclerotic glomeruli compared to normal baseline. It is likely that such additional proteins, perhaps of higher molecular weights, not detected by our LCM and MALDI MS approach, would also be

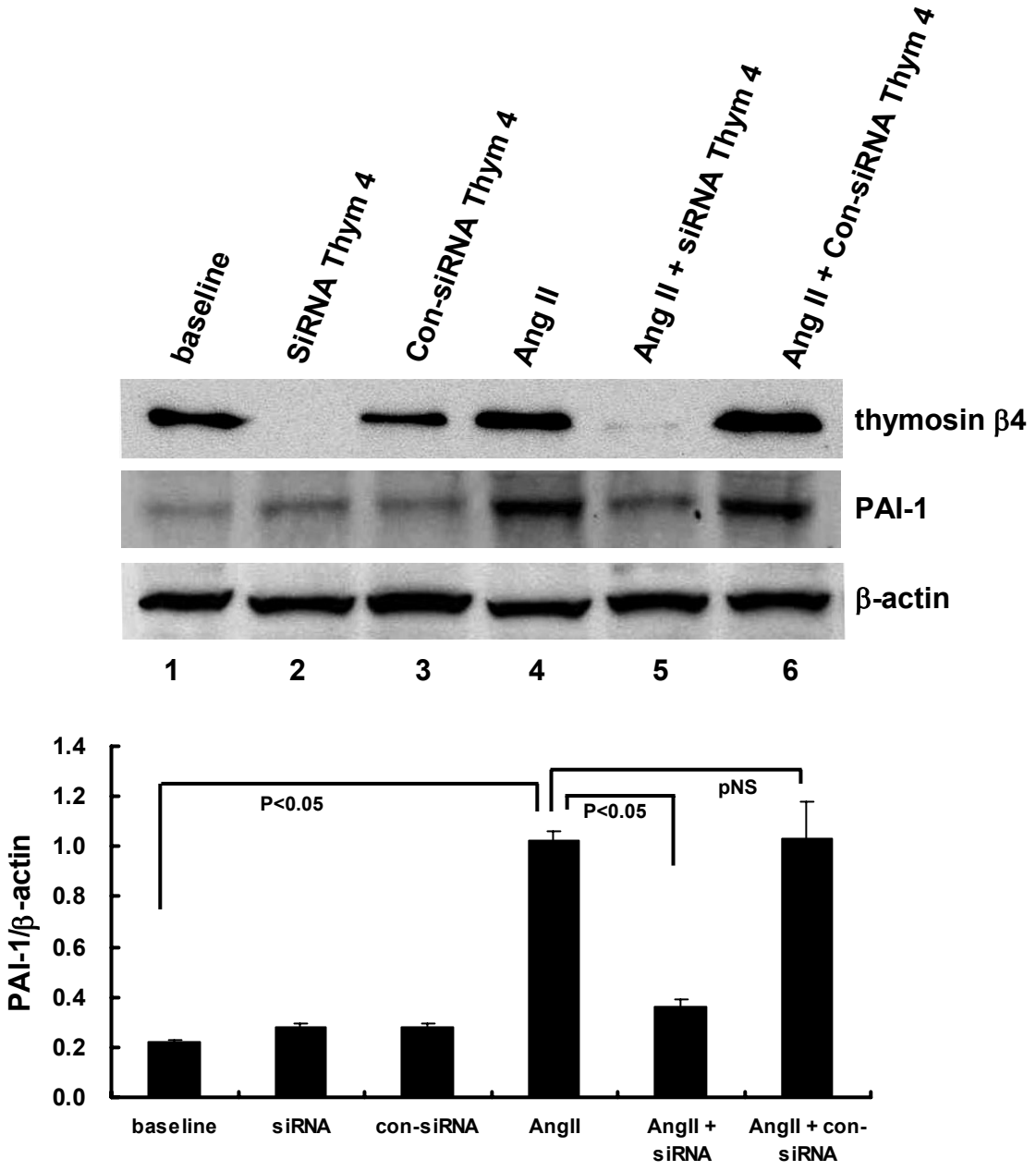


Figure 30: Top: Representative Western blot of replicate experiments of thymosin β4 and Ang II-induced PAI-1. GEN were stimulated with Ang II (10-6 M) for 24 hours, with or without concomitant transfection with siRNA or control siRNA for thymosin β4. Normal GEN at baseline expressed thymosin β4 (lane 1), which was successfully knocked down about 90% using siRNA (lane 2 and 5 vs lane 3). Neither siRNA nor control siRNA affected baseline PAI-1 expression (lane 1 vs lane 2 and lane 3). Ang II dramatically upregulated PAI-1 in normal GEN (lane 4). Transfection with siRNA markedly dampened this Ang II induction of PAI-1 (lane 5). Control siRNA had no effect (lane 6). Bottom: Average PAI-1 protein expression of replicate Western blot experiments, normalized to β-actin.

differentially upregulated in both nonsclerotic and sclerotic glomeruli vs. normal. However, podocyte effacement does not per se correlate with progressive sclerosis. In human disease, both FSGS and minimal change disease (MCD) are characterized by diffuse foot process effacement. However, nonsclerosis develops in MCD, and prognosis is excellent, in contrast to the progressive scarring of FSGS [89]. In small human renal biopsies, it may be difficult to differentiate between MCD vs. FSGS, if the defining segmental sclerotic lesion is not sampled. Differential downregulation of GBM α- and β-dystroglycan by immunostaining in MCD, but not FSGS, has been proposed as a tool for distinction of these two entities [95]. We hypothesize that increased thymosin β4 may be an additional such marker to allow identification of early sclerotic processes.

Our data further indicate that the nonsclerotic glomeruli in FSGS not only have foot process effacement, but also have activation of prosclerotic mechanisms. Our data thus has important implications for therapy, as it suggests that these glomeruli are further activated towards sclerosis than standard light microscopy and electron microscopy would suggest, and more targeted, aggressive therapy might be needed to effect remission or even regression of sclerosis [90].

The identification of the differentially expressed protein markers in the early, prosclerotic stage of progression could advance our understanding of pathomechanisms and possibly direct such therapies. As more 5/6 Nx kidney sample become available, more of the top protein markers could be identified. Currently, with limited tissue left for protein extraction after the LCM experiment, we found by proteomic analysis thymosin β4 expression levels was found to be elevated in sclerotic and nonsclerotic glomeruli vs. baseline. This upregulation of thymosin β4 was further confirmed to be localized to endothelial cells by immunohistochemistry *in vivo* and cell culture analyses. Thymosin β4 is a highly conserved protein which has a wide range of functions. Thymosin β4 is an intracellular G-actin sequestering proteins [96], and plays a role in

wound healing, angiogenesis, and endothelial cell differentiation [97]. Recently, thymosin β4 was reported to increase plasminogen activator inhibitor 1 (PAI-1) expression at both the mRNA and protein levels in endothelial cells [98]. PAI-1 inhibits tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA), preventing the activation of plasminogen to plasmin, which further degrades not only fibrin, but also extracellular matrix (ECM) [99]. PAI-1 also inhibits u-PA-induced matrix metalloprotease-2, thus further inhibiting ECM degradation. PAI-1 is induced by angiotensin II in vivo and in vitro, and is tightly linked to glomerulosclerosis [100]. Conversely, PAI-1 downregulation is linked to amelioration or even regression of glomerulosclerosis [90]. Therefore, upregulation of thymosin β4 could promote the glomerulosclerosis process via upregulation of PAI-1.

To further explore the role of thymosin β4 in this sclerotic pathway, we modulated its expression in cultured endothelial cells. Downregulation of thymosin β4 by RNAi technology decreased angiotensin II-induced PAI-1 expression. These findings imply that thymosin β4, via effects on PAI-1, is not only a marker, but potentially a contributor, to glomerulosclerosis. Thus, thymosin β4 could be another novel target for treatment of glomerulosclerosis.

In conclusion, using laser capture microdissection combined with MALDI mass spectrometry technology, specific proteomic patterns were obtained that accurately classified normal vs. non-sclerotic or sclerotic glomeruli in FSGS. The proteomic pattern of nonsclerotic glomeruli in a fibrosing kidney was found to be more similar to the proteomic pattern of sclerotic glomeruli than to normal glomeruli, suggesting that non-sclerotic glomeruli have early activation of prosclerotic mechanisms. As a discovery tool, our proteomic study further found thymosin β4 to be a key protein marker of glomerulosclerosis and possibly even a contributor to progression by promoting angiotensin-induced PAI-1 expression.

Materials and Methods

Remnant Rat Kidney Model

Adult male Sprague Dawley rats (n=6, 250-300g, Charles River, TN, USA) were studied. Rats were housed under normal conditions with a 12-hour light/dark cycle, at 70°F with 40% humidity and 12-air exchanges/hour and received normal rat chow and water ad lib ("5001" diet, Purina Laboratory Rodent diet, 23.4% protein, 4.5% fat, 6.0% fiber, 0.40% sodium). Rats underwent 5/6 nephrectomy under pentobarbital anesthesia by right unilateral nephrectomy (Nx) and ligation of branches of the left renal artery, producing a total of 5/6 renal ablation. Focal segmental glomerulosclerosis was well developed by 12 weeks after 5/6 Nx, and rats were then sacrificed. Glomerulosclerosis was defined as collapse and/or obliteration of glomerular capillary tuft accompanied by hyaline material and/or increase of matrix [101]. Nephrectomized right kidneys obtained at baseline were used as normal control. A limited number of nonsclerotic glomeruli with normal histology were present in the remnant kidneys.

Laser Capture Microdissection

Protein profiles were obtained directly from laser capture microdissected cells using MALDI MS [51]. Frozen kidney samples from 5/6 Nx or normal baseline were partially embedded in OCT and 5 μm thick cryostat sections were cut. The tissue sections were mounted on regular glass slides and dehydrated as follows: 70% ethanol 30s, 95% ethanol 1 min, 100% ethanol 1 min (2 times), xylene 2.5 min (2 times), and air dried. Rat kidney glomeruli were easily identified and classified as normal vs. sclerotic. Glomeruli were identified by phase microscopy as 1) normal in baseline kidney, and as 2) sclerotic or 3) nonsclerotic in 5/6 Nx. Glomeruli from each category were then microdissected using the Arcturus PixCell II LCM system (Mountain View, CA, USA) with a focused 30 μm laser beam. These three groups of glomeruli were captured separately onto different LCM caps (Mountain View, CA, USA). On average 50 normal

glomeruli were captured from the control right kidney obtained during nephrectomy for each rat. From the remnant kidneys 12 weeks after 5/6 nephrectomy, 30 non-sclerotic and 50 sclerotic glomeruli were obtained from each rat using LCM. Nonsclerotic glomeruli were obtained from 5 of the 6 rats because one rat did not have sufficient nonsclerotic glomeruli present in the ablated kidney. All 6 rats with 5/6 Nx had sufficient sclerotic glomeruli for LCM analysis.

MALDI Mass Spectrometry

Protein expression profiles can be directly obtained from laser microdissected cells using MALDI MS [51]. The LCM thermoplastic films with captured cells were peeled off from the LCM cap using forceps and mounted on a MALDI target plate using a conductive double-sided tape (Digi-Key, Thief River Falls, MN, USA). Sinapinic acid (Sigma, St. Louis, MO, USA) solution (20 mg/ml, 50/50/0.3, v/v/v, acetonitrile/water/TFA,) was microspotted on the captured cells under microscope visualization using pulled fine glass capillaries. Upon solvent dehydration, the matrix and proteins co-crystallized together. MALDI MS analyses were performed in the linear mode under the optimized delayed extraction condition on an Applied Biosystems DE-STR Voyager mass spectrometer (Framingham, MA, USA). The crystals was irradiated by a series of pulsed laser (λ = 337nm) in a high vacuum. The protein analytes were desorbed and ionized from the surface of the tissue forming predominately singly charged protonated ions of the form $[M+H]^+$, where M was the protein molecular weight. The protein ions were accelerated in a constant electric field and subsequently separated in a time-of-flight (TOF) mass analyzer. The mass-to-charge (m/z) ratio of each protein ion was recorded at the detector. Each mass spectrum was obtained averaging signals from 250 consecutive laser shots from five microdissected glomeruli. After internal calibration, the mass spectra were baseline subtracted and normalized using the software developed in our laboratory prior to statistical analysis.

Protein Identification

Remaining frozen rat kidney cortex (114 mg) samples from the same samples investigated by LCM were homogenized in 500 µl of protein extraction buffer using an ice-chilled Duall glass homogenizer. The protein extraction buffer is composed of 0.25 M sucrose (J.T. Baker, Phillipsburg, NJ, USA), 0.01 M Tris-HCl (J.T. Baker) and 0.1mM PMSF (Sigma). The homogenate was centrifuged according to the following sequence: 10 minutes at 680g, 10 minutes at 10,000g, and 1 hour at 100,000 g. The final supernatant was filtered using a Millipore Ultrafree-MS 30,000 NMWL centrifugal filter device (Bedford, MA, USA). A volume of 150 µl filtrate was separated on a Vydac 259VHP5415 polymeric column (Hesperia, CA, USA) at 40°C using a Waters Alliance HPLC system (Milford, MA, USA). Solvent A contained 0.1% TFA (Burdick and Jackson, Muskegon, MI, USA) in water and solvent B contained 0.085% TFA in acetonitrile (Fisher Scientific, Fair Lawn, NJ, USA). A flow rate of 1 ml/min was used with a gradient started from 5% B for 5 min, then in 55 min to 60% B, then in 10 min to 95% B and hold at 95% B for 10 min. The fractions were collected every minute and then completely dried using Thermo Quest Savant Speedvac (Holbrook, NY, USA). Dried HPLC fractions were reconstituted in 10 µl of 5/5/0.3, v/v/v, acetonitrile/water/TFA and analyzed by MALDI-TOF MS.

The fractions containing target protein markers, as identified by statistical analysis (see below) from LCM samples were completely lyophilized again and reconstituted with 10 µl of 0.4M ammonium hydrogen carbonate (Sigma). These fractions were reduced with 5µl of 45mM dithiothreitol (Sigma) in incubation at 60°C for 15 min, followed by alkylation with 5 µl of 100mM iodoacetamide (Sigma) in the dark for 15 min. One microliter of 1 µg/µl sequencing-grade trypsin (Promega, Madison, WI, USA) was added, and the digestion allowed proceeding for 4 hours at 37°C.

The digested fractions were subjected to LC-MS/MS analysis using a ThermoFinnigan LTQ mass spectrometer (San Jose, CA, USA). Two microliters of sample were loaded into a 100

μm i.d. self-packed micro-capillary reverse phase column packed with Monitor C18 -Spherical Silica from Column Engineering Inc (Ontario, CA, USA). The mobile phase A was 0.1% formic acid (EM Science, Darmstadt, Germany) in water and phase B was 0.1% formic acid in acetonitrile. The gradient for mobile phase B started at 0% for 3 min, to 5% in 2 min, to 50% in 45 min, and to 90% in 5 min. The flow rate at the source was 700 nl/min. The fragment ion mass spectra were used to search the National Center for Biotechnology Information (NCBI) rat protein database using the SEQUEST algorithm [27].

Immunohistochemistry

For immunostaining, remnant rat kidney tissue was fixed in 4% paraformaldehyde overnight at 4°C, routinely processed and embedded in paraffin. Four micrometer sections were treated with 3% hydrogen peroxidase for 10 min, Power block (BioGenex Laboratories, San Ramon, CA) for 45 mins, incubated with primary rabbit anti-thymosin β4 antibody (Biodesign, Saco, ME) for 1 hour at 37°C, and rinsed twice with PBS. HRP conjugated-swine ant-rabbit antibody (Dako, Carpenteria, CA) was added and incubated for 45 mins at RT. After rinsing 3 times with PBS, diaminobenzidene was added as a chromagen. Slides were counterstained with hematoxylin.

Infiltrating macrophages were detected with double staining for thymosin β4 and mouse monoclonal antibody to ED1 (BioSource International, Camarillo, CA), a macrophage marker, followed by biotinylated goat anti-mouse IgG (BioGenex) for 30 min and alkaline phosphatase-streptavidin conjugate (BioGenex) for 30 min. Sections were developed with sigma fast red TR/Naphtol AS-MX for 5 min, then counterstained and cover-slipped. Negative controls omitting the primary antibody and using nonspecific immunoglobulin showed no staining. Positive control using rat spleen showed the expected distribution of thymosin β4 [102]. Glomerular endothelial cells and mesangial cells were identified on serial sections by immunostaining with anti-RECA-1

antibody (Abcam, Cambridge, MA) and mouse anti-rat CD90 (Thy-1) antibody (BD Pharmingen, San Diego, CA), respectively.

Cell Culture

Glomerular endothelial cells (GEN), derived from SV40 mice (gift from Dr. Michael Madaio) were grown in 10% fetal bovine serum (FBS) which had been heat-inactivated at 56^{o}C for 1 hour, with DMEM:Ham F12 media (low glucose DMEM, 6mM) in a 3:1 ratio, with L-glytamin 2mM and HEPES 10mM added [103]. The cells were grown at 37^{o}C, with 5% CO_2 under humid conditions in Corning flasks. The cells showed CD31 expression, confirming endothelial cell phenotype. Primary cultures of podocyte were performed as follows: rat kidneys were removed and the renal capsules were stripped using autoclaved surgical instruments. The cortex was isolated and minced into small pieces with razor blade, and glomeruli isolated by sieving (sieve pore size 180 μ X2, 75 μ X1). Glomeruli were then suspended in DMEM/Ham's F-12 (2:1) containing 0.2 um-filtered 3T3-L1 supernatant, 5% heated-inactivated FCS, ITS solution, and 100u/ml penicillin-streptomycin. The cells were then plated onto collagen type I-coated flasks, and incubated at 37^{o}C, room air with 5% CO_2. After 4 days, cell colonies began to sprout around the glomeruli. Cells showed an epithelial morphology with a polyhedral shape when confluency was reached at day 7. The cells were characterized as podocytes by detection of podocyte specific markers, synaptopodin and nephrin, by immunofluorescence staining.

siRNA Design and Transfection

Control siRNAs and siRNAs (antisense and sense strands) for thymosin β4 (Thym) were designed suggested by the manufacture (Invitrogen, San Diego, CA, USA). The sense strand sequences for 4 different siRNAs and scrambled controls were as follows: Thym 1 5'-CCGATATGGCTGAGATCGAGAAATT; Thym 2 5'-GAG AAG CAA GCT GGC GAA TCG TAA T; Thym 3 5'-TCA AAG AAT CAG AAC TAC TGA GCA G; Thym 4 5'-GGG AGA

TGA TGA AAT AGA GAG GAA A; control Thym 1 5'-CCG GGT AAG TCC TAG AGA GAT AAT T; control Thym 2 5'-GAT CCA TGC AGC GTA TCC GAT GAA T; control Thym 3 5'-TCA TAA GAG ACA TCA AGT CGA ACA G; control Thym 4 5'-GGG ATG ATG AAA TAG AGA GGA GAA A. In vitro transfections were performed using Lipofecta^{TM} 2000 Reagent (Invitrogen) according to the manufacturer's instructions. In brief, GEN were seeded on to 6-well plates one day prior to transfection. Transfection with siRNA was done when cells were above 50% confluent. 100 pmol of siRNA were used for 5 x 10^5 cells in 2 ml of medium. Cells were washed 48 hours after transfection. Angiotensin was used to stimulate GEN as a model of sclerosis mechanisms. GEN were stimulated with angiotensin II (Ang II, 10^{-6} M) for 24 hours, with or without concomitant transfection with siRNA or control siRNA. Since all 4 designed siRNAs achieved equal downregulation of thymosin β4, only one set of siRNA and its scrambled control was used for these experiments. Results were compared to normal GEN as baseline control.

Western Blot Analysis

For thymosin β4 Western blot analysis, 100 μg of cell lysate from cultured GEN treated as above were separated by electrophoresis on 4-20% Tris–glycine gel (BioRad). Equal protein loading was confirmed by Coomassie blue staining of duplicate gels after electrophoresis. The gels were incubated for 1 hour in phosphate-buffered saline (PBS) containing 10% glutaraldehyde (Sigma), washed three times for 20 minutes in PBS, and further incubated in a blotting buffer containing 1x Tris-glycine transfer buffer (Invitrogen) and 20% methanol for 30 minutes at room temperature. Proteins were transferred to a nitrocellulose membrane (BioRad) by electrotransfer. The membrane was pre-incubated for 2 hours in PBS containing 5% skim milk and 0.05% Tween 20 (PBS-T), incubated for 1 hour at RT in PBS-T with specific antibody (rabbit polyclonal anti-thymosin β4, 1:1000; Biodesign), washed x5 with PBS-T and incubated with horseradish

peroxidase–conjugated secondary antibody (Amersham Biosciences, Buckinghamshire, UK) for 1 hour at RT. The membranes were washed x5 with PBS-T, and bound antibody was detected with an enhanced chemiluminescence detection kit (Amersham Biosciences). Mouse spleen and muscle tissues were used as positive and negative controls, respectively.

For Western blot analysis of plasminogen activator inhibitor-1 (PAI-1), GEN cell lysate was separated by electrophoresis on 10% Tris–glycine gel (BioRad). After the transfer of protein, the membrane was incubated with antibody specific for sheep anti-mouse PAI-1 antibody (1:250, American Diagnostica Inc. Greenwich, CT). After incubation with horseradish peroxidase (HRP)-labeled anti-sheep IgG secondary antibody (1:1000 dilution in 5% milk TBS-T), the protein bands on Western blots were visualized as above and developed on film. The membranes were restripped for beta actin, used as a housekeeping control protein (Sigma) [94].

Statistical Analysis

The primary objectives of this study were to identify protein patterns that correlate with biological groups and assess the protein patterns closeness among three biological groups. The statistical analyses for identifying a set of proteins expressed differentially between biological groups were focused on the following steps: (1) the selection of important proteins that were differentially expressed among the study groups. The selection was based on Kruskal-Wallis test, Fisher's exact test (dichotomize the expression level as present or not), permutation t-test, Significance Analysis of Microarrays (SAM)[104], Weighted Gene Analysis (WGA)[105], and the modified info score method [106, 107]. The proteins were ranked based on the results from all these six methods. (2) The Weighted Flexible Compound Covariate Method (WFCCM)[22, 108-110] was employed in the class-prediction model based on the ranked proteins to verify whether the proteomic patterns could be used to classify study samples into two classes according to the chosen parameter, e.g., normal vs. non-sclerotic. The WFCCM method was designed to combine

the most significant proteins associated with the biological status from each analysis method, and it reduced the dimensionality of the problem using a new covariate obtained as a weighted sum of the most important predictors. We estimated the misclassification rate using leave-one-out cross-validated class prediction method based on the WFCCM. (3) The agglomerative hierarchical clustering algorithm was applied to investigate the pattern among the statistically significant discriminator proteins as well as the biologic status using M. Eisen's software [65].

The statistical analyses for assessing the profile closeness between biological groups were focused on the following steps: First, the profile difference between two biological groups was measured using the Weighted Gene Analysis method for each individual protein. This method is based on the between and within group Euclidean distance. After the profile difference was estimated, the assessment of the overall closeness among three biological groups was completed using the restricted/residual maximal likelihood (REML)-based mixed effect model to test the significance of the closeness among three groups.

The comparison of quantitative analysis of thymosin β4 among three classes of glomeruli was completed using the General Linear Model (Analysis of Variance) to test the mean difference. The p-values as well as the 95% Confidence Interval (CI) were reported. All the tests of significance were two-sided, and the differences were considered statistically significant when p-value < 0.05

Acknowledgement:

The 5/6 nephrectomy rat model, thymosin β4 immunohistochemistry and in vitro analyses were performed by Dr. Fogo's laboratory. The statistical analysis was performed by Dr. Shyr's laboratory.

CHAPTER V

PROTEOMIC PATTERNS OF TUMOR SUBSETS IN HUMAN BREAST CANCER

Introduction

Breast cancer is the second leading cause of cancer-related deaths in women of the western world. Early-stage breast cancer diagnosis is a major reason for improvement in the breast cancer mortality rate. Development of low grade invasive mammary carcinoma is likely a sequential progression starting with alterations in normal mammary epithelium, resulting in atypical ductal hyperplasia (ADH), which may progress to low grade ductal carcinoma *in situ* (DCIS), and culminating in invasive mammary carcinoma (IMC). High grade DCIS is accepted as the precursor of high grade invasive mammary carcinoma; however, the intermediary preneoplastic lesions between normal mammary epithelium and high grade DCIS are essentially unknown. Thus mammary tumorigenesis is postulated to occur along at least two pathways, which may be largely independent. Invasive mammary carcinomas of any grade may also acquire the capacity for distant metastasis to other organs [111].

Normal mammary epithelium is composed of a single layer of luminal epithelial cells which sit above a layer of myoepithelial cells and line normal mammary ducts and lobular units (Figure 31A). By definition, DCIS consists of a monotonous tumor cell population completely filling but confined to basement membrane bound ducts and lobular units and measuring at least 3.0 mm in size. Practically speaking, minimal DCIS are usually 5-8 mm and low grade, where as high grade DCIS are often 2.0-5.0 cm or more (Figure 31B). When mammary carcinoma has achieved the capacity for invasion, tumor cells infiltrate the stroma compartment of the breast.

The DCIS is accepted as the precursor of IMC (figure 31C), based on its common concurrence with IMC and its high rate of recurrence as an invasive tumor at the original site

when incompletely excised. The transition from DCIS to IMC is a critical event in the development of mammary neoplasia. In women with untreated DCIS, death is a much later event than in women with invasive cancer. The natural history of DCIS also demonstrates that in some cases the capacity for invasion is never achieved, because when DCIS is present alone, it is completely curable by local excision [112]. By defining the protein alterations which can distinguish DCIS (lesions which by definition lack the capacity for invasion) from IMC (lesions which by definition are invasive), protein markers characteristic of invasion can be elucidated.

The heterogeneous nature of most tissues poses great challenges for traditional proteomic approaches. Mammary ducts and lobular units represent a relatively small proportion of normal mammary tissue, which is dominantly adipose tissue and fibrous stroma. Similarly, nests of infiltrating mammary carcinoma cells are usually mixed with desmoplastic stroma and inflammatory cells. Direct tissue protein profiling using combined laser capture microdissection and MALDI mass spectrometry is a powerful technology for the rapid and precise discovery of protein markers in complex tissues. This combination of techniques permits simultaneous examination of hundreds of proteins from a nearly pure cell population removed from a heterogeneous sample. Direct analysis of proteomic patterns among IMC, DCIS, and normal mammary epithelium by MS has the potential to rapidly address changes that control the heterogeneous biology of breast cancer. Comparing the protein profiles of normal mammary epithelium, DCIS, and IMC should identify unique protein signatures characterizing each of these classes. Subsequent identification of the component proteins should lead to the discovery of novel targets for new preventative and therapeutic strategies as well as enhance our understanding of mammary tumorigenesis.

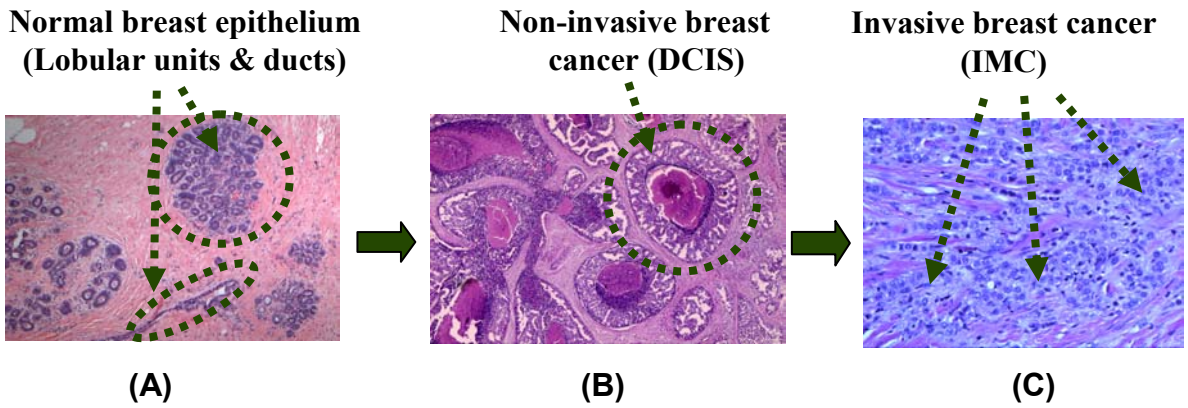


Figure 31. Breast cancer progression from normal lobular units & ducts (A) to ductal carcinoma in situ (DCIS) (B) and to invasive mammary carcinoma (C). The heterogeneous nature of the breast cancer tissue requires the microdissection of specific cells of interest for comparative proteomic analysis.

Specific Aim

To differentiate human breast tumor from normal breast tissue based on their proteomic patterns and begin to distinguish in situ from invasive cancer at the protein level to aid breast cancer diagnosis.

Results

Reproducibility and Precision of Protein Profiling

The tissue samples were collected from four different centers by way of the Cooperative Human Tissue Network (CHTN). First, we wanted to determine if slight but inevitable differences in specimen procurement and handling would affect the protein profiling. We performed a data agreement analysis based on four examples of normal mammary tissue procured from each of the four centers. For each individual specimen included in the data agreement analysis, three separate protein profiles were generated from microdissected normal mammary epithelium and designated as A, B, and C, for a total of 48 representative MALDI-MS profiles. After internal calibration, baseline subtraction and intensity normalization, we compared all peaks in these three independently obtained spectra by pairwise comparison and determined the percentage intra-sample agreement. Data agreement analyses demonstrated 84% individual intra-sample agreement with an 86.7% overall intra-sample agreement rate. Similarly, we wanted to compare the intra-sample variability across collection centers. These analyses showed intra-sample agreement across centers to be essentially the same as the overall intra agreement rate. Variance component analysis based on a mixed effect model examining the intra-sample variation with respect to total variation demonstrated an intra-class correlation coefficient (ICC = inter-sample variation/inter-sample variation + intra-sample variation) of >70% in the 2500 Da to 25 kDa range.

Protein expression profiles of normal mammary tissue and breast tumor

Protein expression profiles were generated from 88 examples of normal mammary epithelium (obtained from reduction mammoplasty specimens), 30 examples of DCIS and 92 examples of IMC. These specimens were from 196 women. In 14 cases both IMC and DCIS were microdissected from the same biopsy. Thirteen DCIS cases were obtained from women with concurrent IMC but their IMC was unavailable for dissection, while 3 cases were from women with DCIS only. All these specimens were microdissected using LCM by a board-certified pathologist (Melinda Sanders, M.D.). For each patient, approximately 500-1000 cells were captured and an average of three mass spectra was obtained. The multiple spectra from each patient were averaged as one spectrum. Approximately 200 peaks were detected per spectrum and more than 1600 signals were detected across all spectra obtained from the 196 patients. Representative spectra from three different tissue classes are shown in Figure 32, displaying the protein profiles of microdissected normal mammary epithelium, DCIS and IMC from the m/z range of 3,200 to 70,000. Asterisks mark the differentially expressed protein peaks between the normal, DCIS and IMC samples, showing that there are detectable differences in protein expression between the different tissue classes.

Normal Mammary Tissue, DCIS and IMC Classification

Using the weighted flexible compound covariate method (WFCCM) and hierarchical clustering analyses, the following proteomic pattern comparisons were performed: 1) Breast tumor (IMC and DCIS) vs. normal mammary epithelium; 2) IMC vs. normal mammary epithelium; 3) DCIS vs. normal mammary epithelium; 4) IMC vs. DCIS.

First, for the breast tumor vs. normal mammary epithelium comparison, 92 IMC and 30 DCIS cases were combined into a single group as mammary carcinoma and compared with the 88

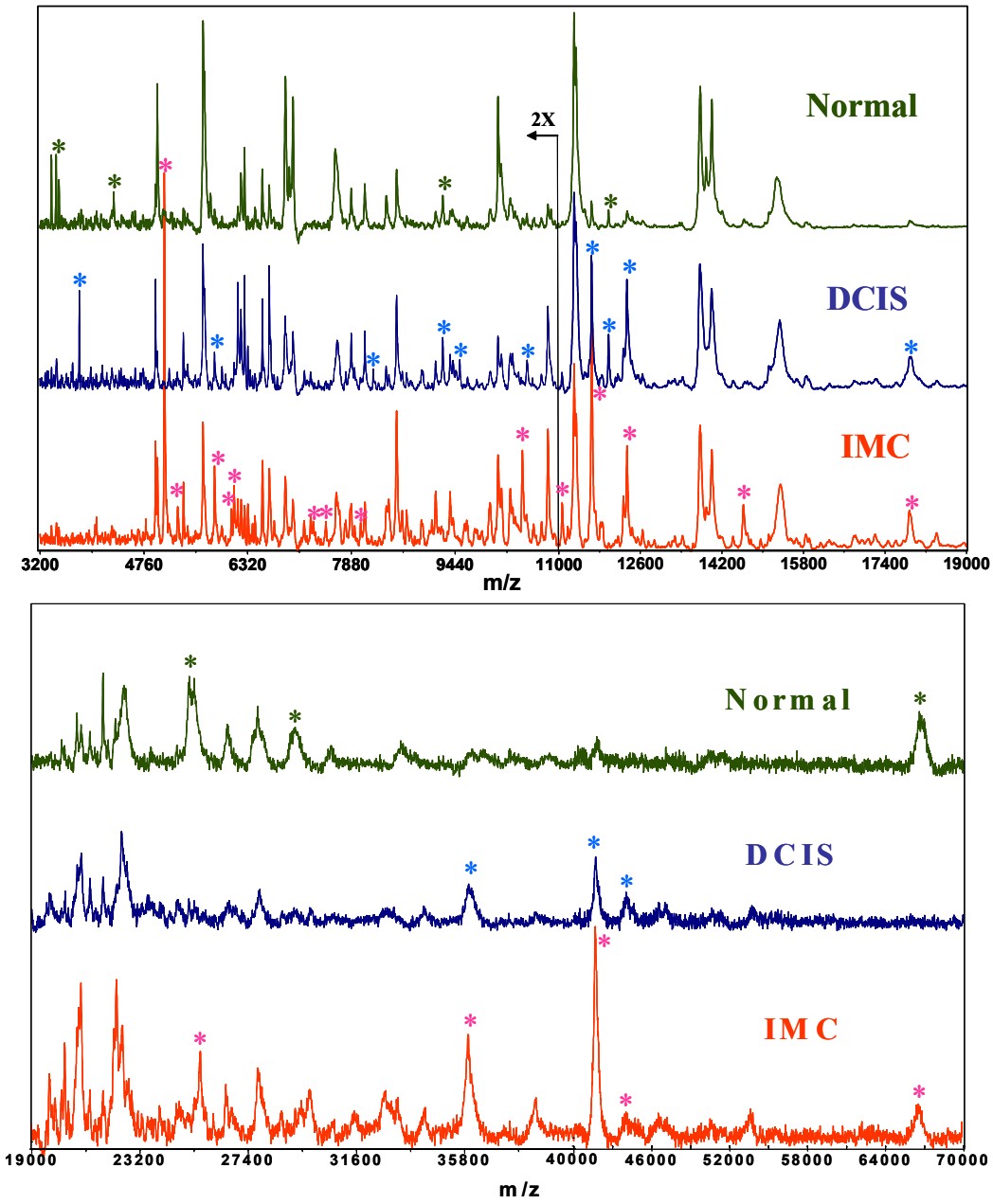


Figure 32. Differences in protein expression patterns among patients with normal mammary epithelium, ductal carcinoma in situ (DCIS) and invasive mammary carcinoma (IMC) can be successfully detected. Three mass spectra from each sample type were averaged to produce the profiles. Asterisks (*) mark the proteins differentially expressed among IMC, DCIS and normal mammary epithelium.

examples of normal mammary epithelium. The 210 samples were equally but randomly divided into training and testing cohorts, with 61 tumors and 44 normal samples in each cohort. The tumor group in each cohort consisted of 15 cases of DCIS and 46 cases of IMC. When IMC and DCIS were dissected from the same specimen, they were assigned to different cohorts. Figure 33 shows the remarkably high accuracy with which the class prediction model can distinguish breast tumor (IMC & DCIS) from normal mammary epithelium based on increasing numbers of discriminatory peaks. The x-axis represents the number of discriminator peaks and the y-axis represents the classification accuracy. Using the top 50 differentially expressed peaks, the class prediction model classified tumors (IMC&DCIS) vs. normal mammary epithelium with 95% accuracy in the training cohort and 88% accuracy in the testing cohort.

Hierarchical clustering analyses of the training cohort showed accurate segregation of the tumor vs. normal samples to opposite ends of the dendrogram and visualized the proteomic pattern distribution (figure 34). IMC perfectly segregated from normal breast. DCIS was interspersed among IMC but tended to cluster at one end of the tumor bracket. IMC clustered roughly by grade, but interestingly the low grade cancers clustered in the center of the dendrogram. One of the two misclassified DCIS could be explained by a technically difficult dissection which may have been contaminated by inflammatory cells and stroma. The other example of DCIS had been poorly visualized at the time of dissection and was determined not to represent DCIS. One of the misclassified normal specimens was later determined to have been normal tissue adjacent to tumor.

Next, the 92 IMC and the 88 examples of normal mammary epithelium were compared. Those 180 cases were from different patients and randomly divided into training and testing cohorts, with 46 IMC and 44 normal samples in each set. The IMC vs. normal mammary epithelium were distinguished with remarkably high classification accuracy by the class prediction model based on increasing numbers of discriminator peaks. Using the top 50

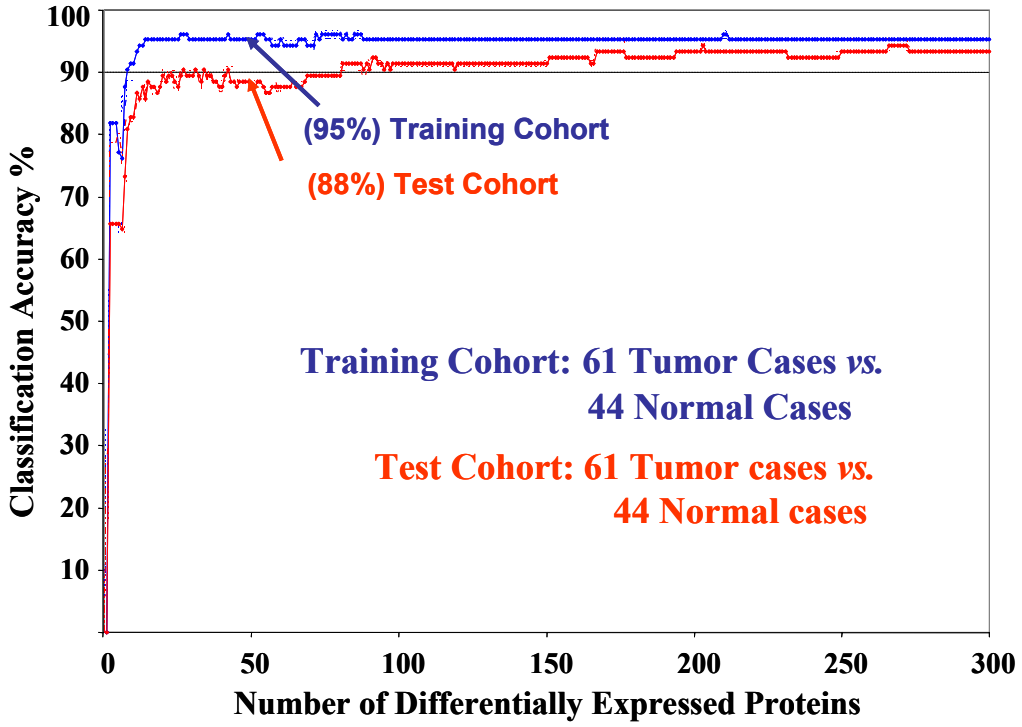


Figure 33. Classification accuracy curves for the training and test cohorts after application of the class prediction model. Normal consists of 44 examples of normal mammary epithelium from reduction mammoplasty specimens in both the training and test cohorts. Using the top 50 differentially expressed peaks, the class prediction model classified tumors (IMC&DCIS) vs. RM with 95% accuracy in the training cohort and 88% accuracy in the testing cohort.

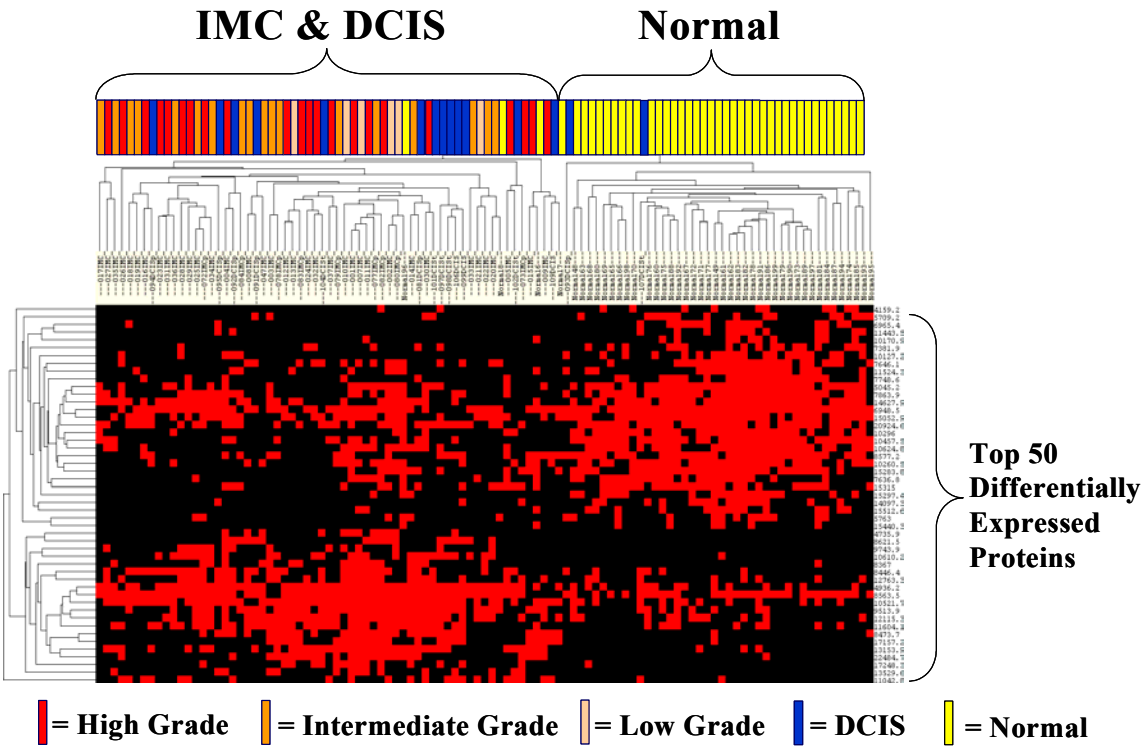


Figure 34. Hierarchical cluster analysis of the 61 breast tumors and 44 normal breast specimens in the training cohort according to the protein expression patterns of 50 MS signals. The cluster at the top shows the similarity in protein expression profiles of the samples. Substantially raised (red) expression of the proteins is noted in individual tumor and normal breast samples. Only 3 normal and 2 tumor cases out of the 105 cases were mis-classified.

differentially expressed peaks, the class prediction model classified IMC vs. normal with 96% accuracy in the training cohort and 90% accuracy in the blinded testing cohort (Figure 35). Subsequent hierarchical clustering analysis showed accurate segregation of the tumor vs. normal samples to opposite ends of the dendrogram using the top 50 protein peaks (Figure 36). From the 90 samples analyzed, only 1 IMC and 2 normal samples were misclassified.

Third, the DCIS and normal mammary epithelium were compared to elucidate patterns of protein expression that are characteristic of DCIS. A total of 30 DCIS and 88 examples of mammary epithelium cases were compared. Because of the relatively small number of DCIS cases, only a supervised analysis was performed and all cases were assigned to a training cohort. Using the leave-one-out cross validation method, 96% classification accuracy was archived using the top 44 protein markers (Figure 37). The hierarchical clustering analysis showed accurate segregation of the DCIS vs. normal samples to opposite ends of the dendrogram using the top 100 protein markers (Figure 38). From the 118 sample analyzed, only 2 DCIS and 5 normal samples were misclassified.

Fourth, 92 IMC and 30 DCIS cases were compared and proteomic patterns associated with the transition from *in situ* tumor to invasive tumor were analyzed. All the cases were assigned to the training cohort. Using the top 72 differentially expressed peaks, the class prediction model classified DCIS from IMC with 80% classification accuracy in the training cohort. When the number of discriminator peaks was increased to 260, the classification accuracy was 93% (Figure 39). The relatively large number of peaks required to accurately distinguish DCIS from IMC indicate, as would be expected, that the protein expression patterns of these two classes are very similar in contrast to the proteomic patterns of normal mammary epithelium vs. tumor. In addition, both the DCIS group and the IMC groups are somewhat heterogeneous in grade. The inclusion of different grades adds complexity to the classification strategy, which also explains why a large number of peaks are required to distinguish the 2 classes. The hierarchical

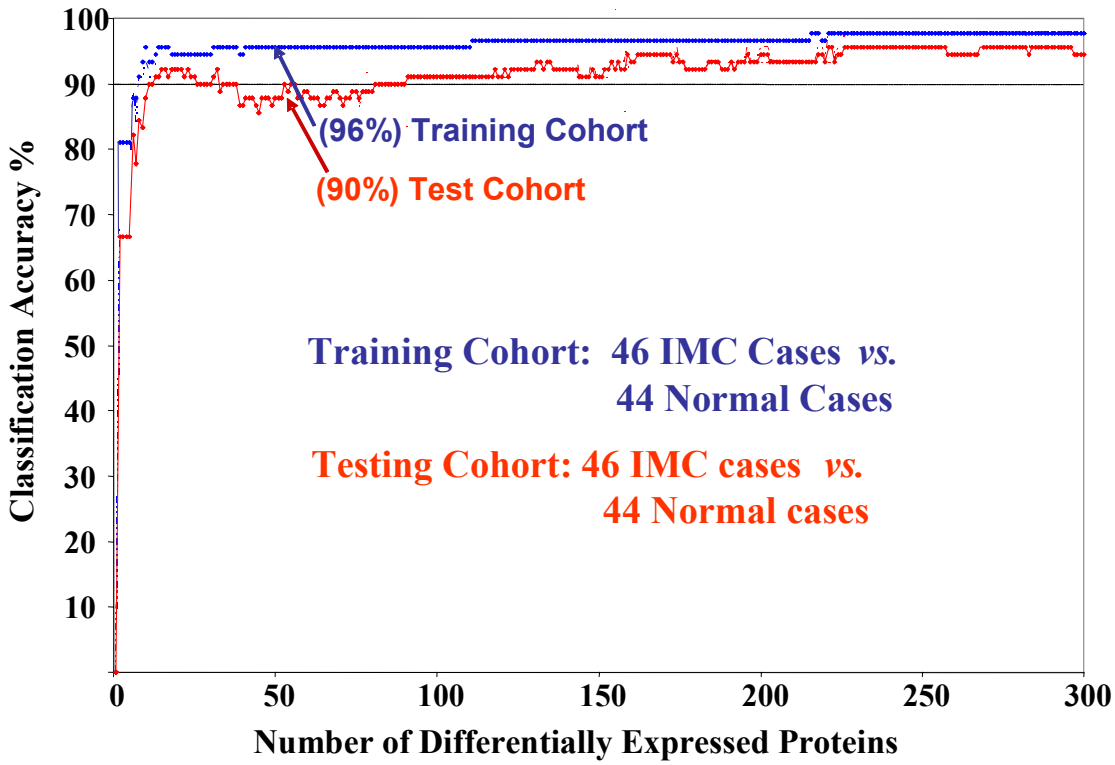


Figure 35. Classification accuracy curves for the training and test cohorts after the application of the class prediction model. Total of 44 normal mammary epithelium and 46 invasive mammary carcinoma are in both training and test cohorts. Using the top 50 differentially expressed peaks, the class prediction model classified tumors IMC vs. Normal with 96% accuracy in the training cohort and 90% accuracy in the testing cohort.

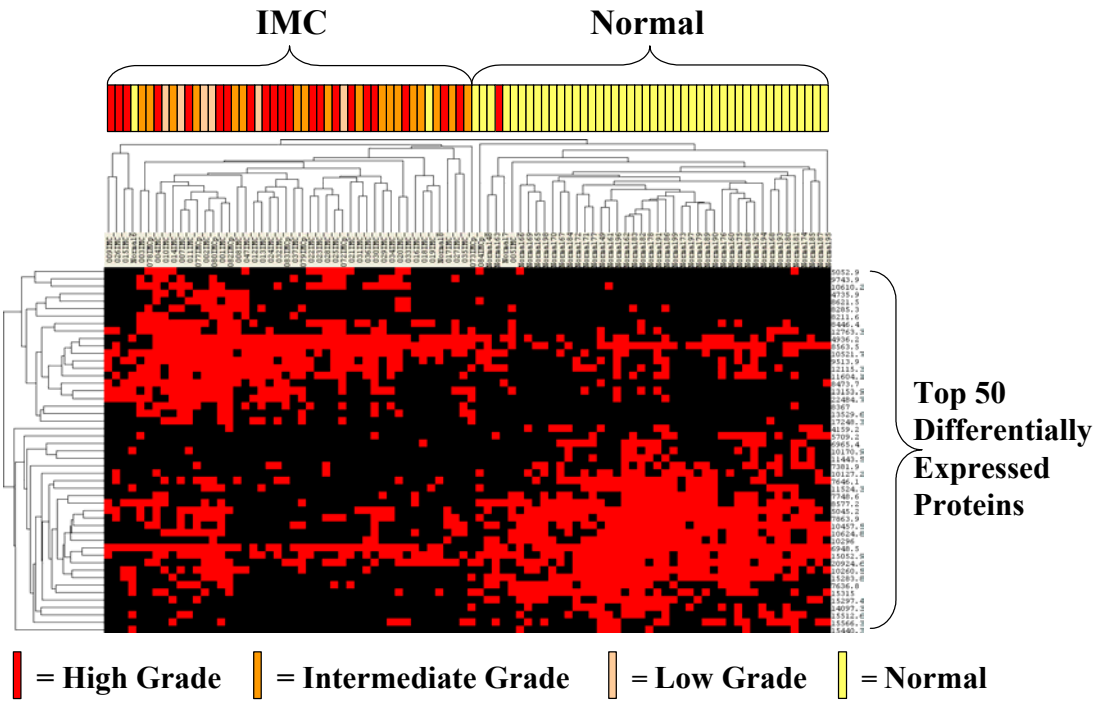


Figure 36. Hierarchical cluster analysis of the 46 invasive mammary carcinomas and 44 normal breast specimens in the training cohort according to the protein expression patterns of 50 MS signals. The cluster at the top shows the similarity in protein expression profiles of the samples. Substantially raised (red) expression of the proteins is noted in individual tumor and normal breast samples. Only 2 normal and 1 tumor cases out of the 90 cases were misclassified.

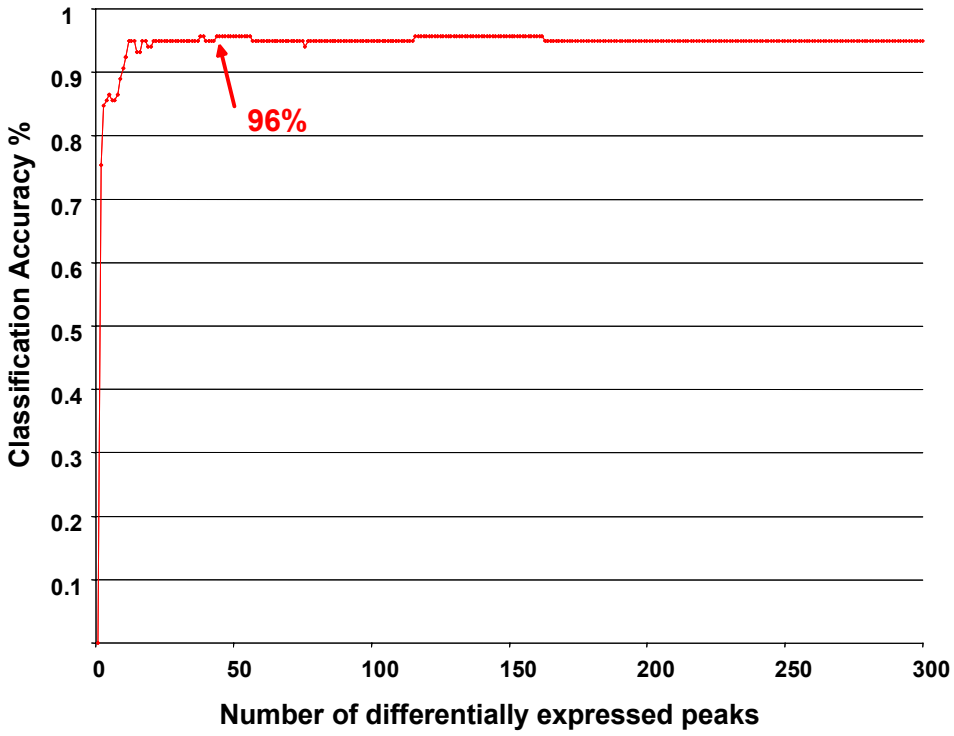


Figure 37. Classification accuracy curves for the classification of normal vs. ductal carcinoma in situ. Total of 88 normal mammary epithelium and 30 DCIS were compared. Using the top 50 differentially expressed peaks, the class prediction model classified tumors DCIS vs. Normal with 96% accuracy using leave-one-out cross validation method.

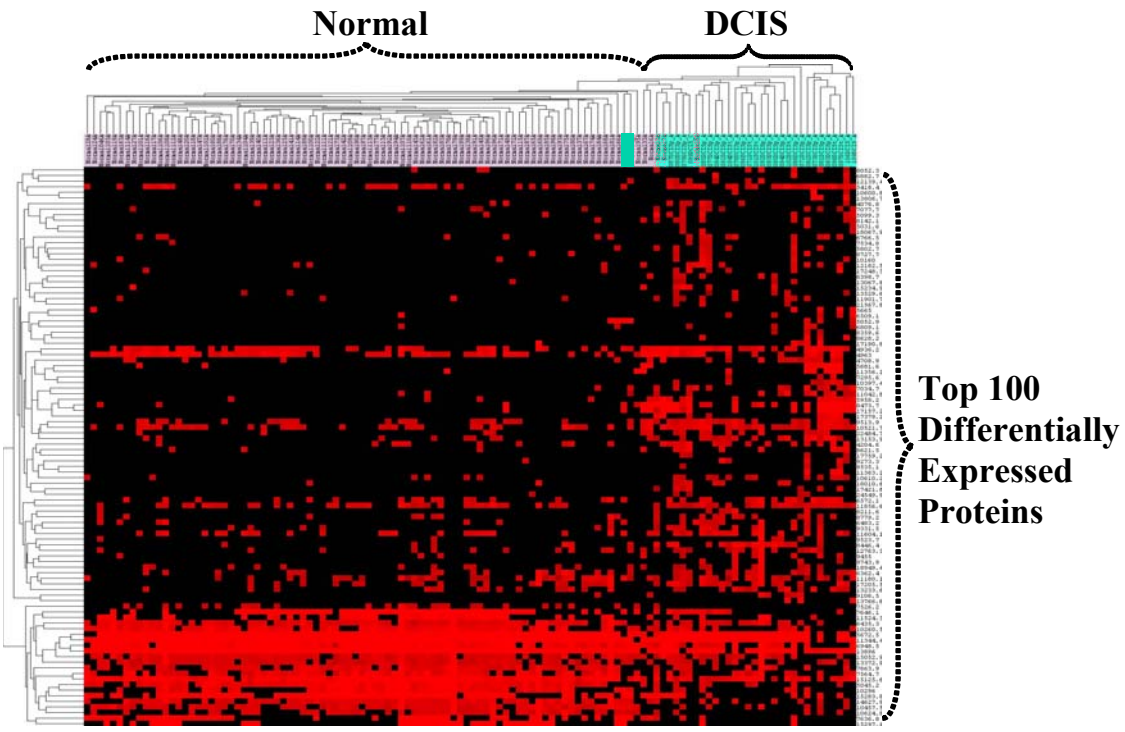


Figure 38. Hierarchical cluster analysis of the 30 DCIS and 88 normal breast specimens in the training cohort according to the protein expression patterns of 50 MS signals. The cluster at the top shows the similarity in protein expression profiles of the samples. Substantially raised (red) expression of the proteins is noted in individual tumor and normal breast samples. Only 5 normal and 2 DCIS cases were misclassified.

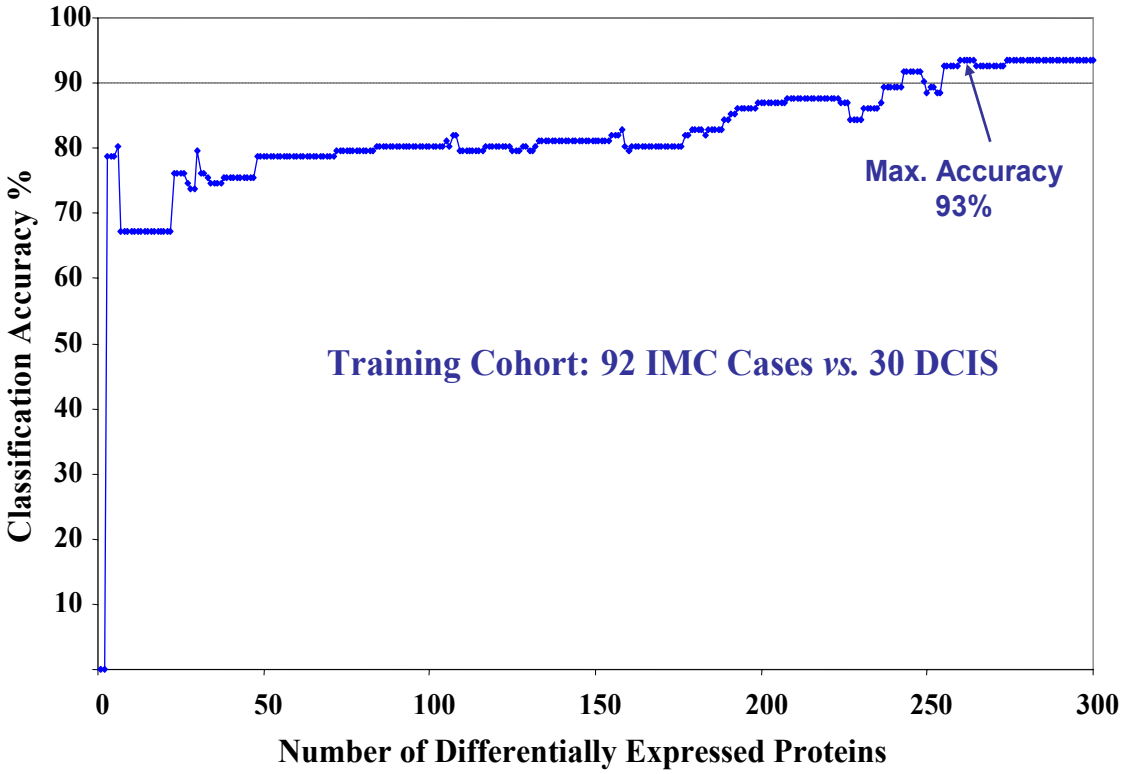


Figure 39. Classification accuracy curves for the classification of IMC vs. DCIS. Total of 92 IMC and 30 DCIS were compared. The classification accuracy increased as the number of protein markers increases. Using the 255 differentially expressed peaks, the class prediction model classified IMC vs. DCIS with 93% accuracy.

clustering analysis showed 6 DCIS cases were misclassified (Figure 40). These data are in accord with the WFCCM result (Fig 40) that the proteomic patterns of DCIS and IMC share a larger number of peaks with each other than either share with normal mammary epithelium.

Protein Marker Identification

The identification of statistically significant protein markers was pursued to explore the biological functions of the differentially expressed protein markers. A schematic representation of the protein marker identification strategy is illustrated in figure 20. Briefly, human breast tumor tissue was homogenized and the cytosolic fraction from the differential centrifugation was further fractionated using reverse phase HPLC. The UV trace at 214 nm recorded the chromatogram (Figure 41A). Each HPLC fraction was measured using MALDI MS and the fractions containing the proteins with the target m/z values were chosen for trypsin digestion.

One of the protein markers which showed greater expression in IMC compared to normal breast, eluted at 38 min in the chromatogram (Figure 42B). The subsequent amino acid sequence analyses of the tryptic peptides were performed using capillary LC-MS/MS and the results were analyzed using the SEQUEST algorithm. The MS/MS spectrum on figure 41C confirmed the identity of the tryptic peptide $[IQYQLVDISQDNALR]_{32\text{-}46}$ from the SH3 domain-binding glutamic acid-rich-like protein 3 (SH3BGRL3). The theoretical molecular weight of SH3BGRL3 is 10348.5 Da, which matches well with the measured protein molecular weight 10347 ± 2 Da when taking into consideration the plausible removal of the initial methionine and subsequent N-terminal acetylation. Glutaredoxin was similarly identified as a protein showing greater expression in the normal breast tissue compared to tumor. The theoretical molecular weight of Glutaredoxin is 11686.6 Da, which matches well with the observed protein molecular weight 11685 ± 2 Da. One tryptic peptide $[DCIGGCSDLVSLQQSGELLTR]_{77\text{-}97}$ was found for this

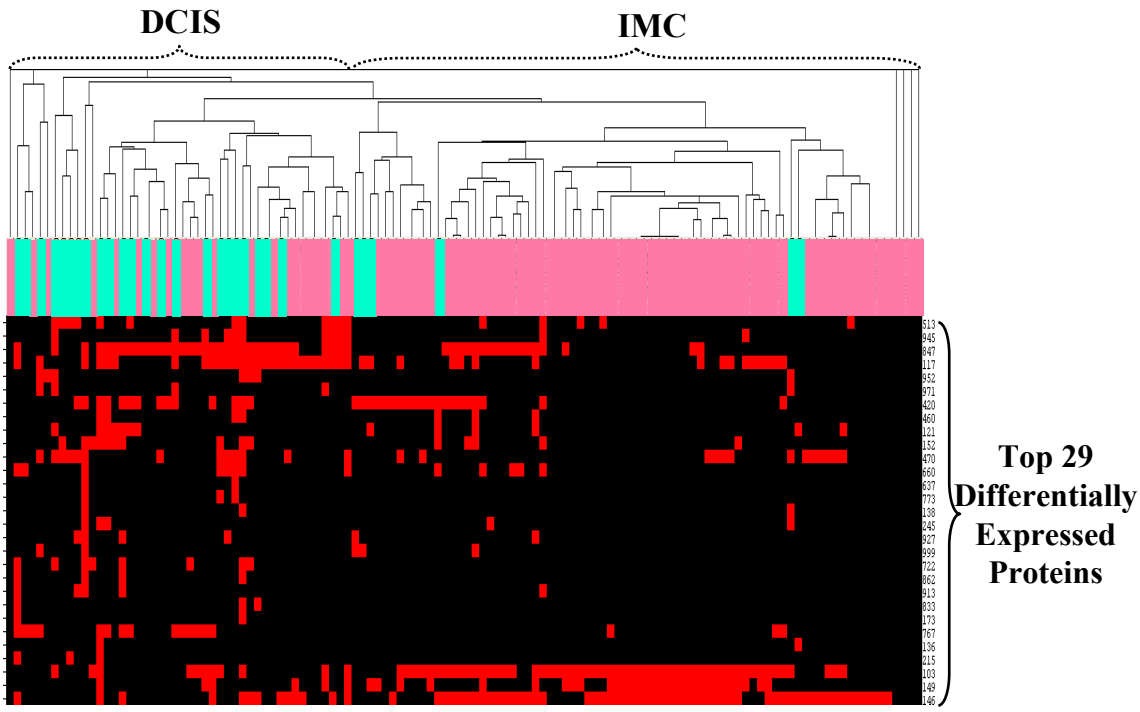


Figure 40. Hierarchical cluster analysis of the 30 DCIS and 88 normal breast specimens in the training cohort according to the protein expression patterns of 29 MS signals. 21 IMC and 6 DCIS cases were misclassified to different clusters.

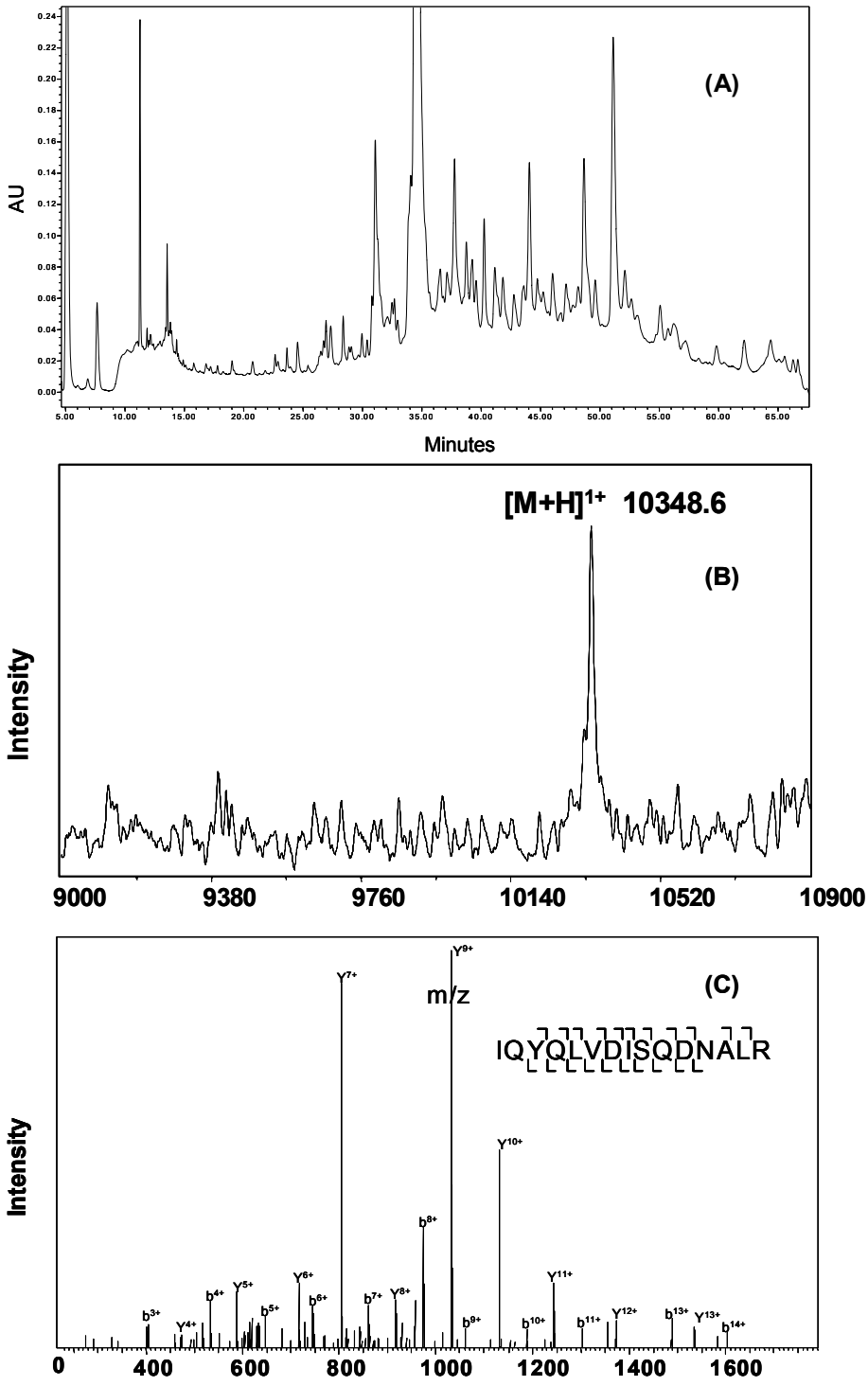


Figure 41. The identification of SH3 domain-binding glutamic acid-rich-like protein 3, a protein marker of invasive mammary carcinoma. A) The UV chromatograph of the HPLC separation of breast tumor cell lysate. B) The protein of 10346 Da, a protein marker of interest, was found in one HPLC fraction. C) The MS/MS spectra were consistent with the tryptic peptides of the target protein.

protein. Thymosin β4 (4964 Da), Thymosin β10 (4937 Da) and ubiquitin (8564 Da) were also identified as highly statistically significant discriminators for IMC vs. normal breast.

Discussion

In this study, protein profiles were obtained from a total of 210 human breast carcinomas and normal breast tissue using LCM combined with MALDI MS. Statistically significant proteomic patterns that accurately distinguished normal mammary epithelium from DCIS and IMC were obtained. A set of the statistically significant protein markers were identified using tandem mass spectrometry.

Selecting the proteomic patterns that can accurately distinguish mammary carcinoma (IMC & DCIS) from normal mammary epithelium was our primary goal for this study. Using the top 50 statistically significant protein discriminators, mammary tumor could be distinguished from normal mammary tissue with greater than 95% accuracy in the training and 88% accuracy in the blinded testing cohorts. We further specifically compared the protein profiles of invasive mammary carcinoma with those of normal mammary epithelium. These two classes of tissues were distinguished with 96% accuracy in the training cohort and 90% in the blinded testing cohort. The proteomic patterns of normal mammary epithelium and DCIS were compared to elucidate protein markers specific to DCIS suggesting that they may be associated with neoplastic transformation. Using the top 47 protein markers, DCIS could be distinguished from normal mammary epithelium with greater than 96% accuracy. Finally, we compared the protein profile of DCIS to IMC to identify expression differences that are associated with the capacity for invasion. Using 255 selected peaks, we distinguished DCIS from IMC with 93% accuracy. The large number of protein biomarkers needed to achieve high classification accuracy suggests the differences in protein expression between the classes are few.

Identification of differentially expressed protein markers selected from the analyses may provide new insights into mechanisms of breast tumorigenesis. Identification of thymosin β4 and thymosin β10 were in agreement with previous reports. Increased expression of beta-thymosins at the RNA or protein level have been reported in lung cancer [22], breast cancer cell lines [113], and colon cancer cell lines [114]. Recent data suggests that thymosin β4, an actin-sequestering protein, may play an important role in tumor metastasis and angiogenesis [115]. Cha et al. showed that overexpression of thymosin β4 is associated with an increase in the expression of vascular endothelial growth factor (VEGF), a known angiogenic factor [116]. In our previous study of thymosin β4's role in glomerulosclerosis, we found that the inhibition of thymosin β4 dramatically decreased the expression level of plasminogen activation inhibitor-1 (PAI-1) via an angiotensin dependent manner. Upregulation of thymosin β4 also increases PAI-1 expression in endothelial cells [98]. Interestingly, high expression of PAI-1 in breast cancer is strongly associated with aggressiveness and poor outcome [117, 118]. Our results thus suggest that thymosin β4 may contribute to breast tumor angiogenesis via the plasminogen activator-plasmin system. Glutaredoxin (GRX) and SH3 domain-binding glutamic acid-rich-like protein 3 (SH3BGRL3), both involved in the regulation of protein redox reactions, were identified as differentially expressed between normal mammary epithelium and mammary tumor cells. GRX belongs to the ubiquitous glutathione (GSH)-dependent oxidoreductase family, which catalyzes the reduction of protein-glutathionyl-mixed disulfides. GRX has various cellular functions, including functioning as a redox sensor via interaction with the apoptosis signal-regulating kinase 1 [119], protection of cells against apoptosis [120], activation of ribonucleotide reductase [121] and regulation of different transcription factors [122]. SH3BGRL3 shows significant sequence homology with GRX and its overall structural fold matches with the thioredoxin fold [123]. SH3BGRL3 is suggested to function as an endogenous modulator of GRX biological activities by competitive binding to yet unknown target proteins.

As larger numbers of tumor samples become available, the study of subpopulations within the invasive cancer group will be possible. First, the presence of lymph node metastases at the time of the primary diagnosis is a significant predictor of relapse and poor outcome. Comparing the protein profiles of lymph node positive and lymph node negative cancers may identify proteins associated with the capacity for metastasis. Such markers would have significant clinical value as they may identify tumors programmed for metastasis before they have spread to lymph nodes. A previous proteomic study of lung cancer found two protein markers which were highly predictive of lymph node involvement [22].

Studying disease outcome in breast cancer is more difficult than other high life-threatening cancers, such as lung cancer, because relapse-free survival can be prolonged, especially for low grade breast cancers. Where as most patients with lung cancer relapse within 6 months to 1 year, breast cancer patients relapse and survival can be prolonged over 15-20 years. Obtaining follow up information over such time periods is challenging and costly. Performing studies of this kind in a timely fashion require the availability of tissue collected decades previously. Future goals of this project are to develop collaborations with other institutions that have tissue databases with follow up.

In summary, using the combination of LCM and MALDI MS, differentially expressed proteomic patterns among normal mammary epithelium and breast tumors were found. This comparative approach to defining these molecular alterations is expected to help identify mechanisms of mammary transformation which may lead to novel strategies for the prevention and treatment of breast cancer. Differentially expressed proteins among benign and malignant epithelium reflect the critical molecular alterations leading to breast cancer development and progression. Identification of these unique protein signatures and subsequent identification of their component proteins may further identify novel tumor specific markers and potential drug targets.

Materials and Methods

Sample collection, tissue microdissection and MALDI MS analysis

Primary breast tumors and normal mammary epithelium from reduction mammoplasty (RM) specimen were obtained from the Cooperative Human Tissue Network (CHTN). All tissues were snap-frozen in liquid nitrogen immediately following removal from the patient. An H&E stained section of each tissue sample was examined to select appropriate areas for LCM. The specific cells of interest were microdissected by a board-certified pathologist (Melinda Sanders, M.D.).

Choosing a reliable normal control was a critical component of this analysis. We decided to use normal mammary epithelium from reduction mammoplasty specimens as our normal population for three reasons. First, these samples are more likely to represent truly normal epithelium as these women do not have breast carcinoma or have a history of breast carcinoma. Second, it has been shown that histologically normal appearing breast tissue adjacent to tumor but not distant to the tumor can have genetic abnormalities identical to those seen in the corresponding tumor [124]. These results may truly represent early stages of neoplastic transformation in apparently normal tissue but is more likely Pagetoid spread, the spread of neoplastic cells in a single cell fashion (a well-known phenomenon to practicing pathologists), into adjacent normal ducts and lobular units. Third, the consistent availability of large numbers of reduction mammoplasty specimens from all the procurement centers makes this population a robust reference group.

MALDI-TOF MS was performed directly on the LCM acquired cells using the procedure described in Chapter II. Depending on the size of the lesion, 400 to 2000 cells were obtained from each sample. Mass spectra were acquired in the liner positive ion mode under optimized delayed extraction conditions using a Voyager DE-STR MALDI mass spectrometer with a 337 nm nitrogen laser (Applied Biosystems). The instrument settings were as follows: accelerating

voltage 25 kV, 91% grid voltage, 150 ns delay time, and 0.05% guide wire voltage. A total of 750 laser shots were averaged to create a single spectrum from the captured cells. The m/z range of the mass spectra was from 2,000 to 70,000.

Data processing and Statistical Analysis

The spectra were internally calibrated using previously identified proteins that are commonly present in the spectra, such as ubiquitin (8565.8 Da), 40 ribosomal S30 protein (6648.8 Da), macrophage inhibitory factor (12345.5 Da) and calmodulin (16792.5 Da). Mass spectra were then baseline corrected, smoothed, and normalized and then combined to create a single spectrum for each sample type. Mass windows (bins) which can identify the same peaks across multiple spectra were generated. Peaks were aligned across samples and these optimal bins were used to define individual protein/peptide peaks within a large data set for statistical analysis. Use of this binning technique is preferable to defining variables at each individual mass unit because it accounts for nonlinear shifts of the mass-to-charge (m/z) values for a given protein between samples.

The statistical analyses for these comparisons involved the following steps: (1) selection of differentially expressed peaks among specimen classes which served as statistically significant discriminators of the compared groups, (2) generation of class prediction models based upon the Weighted Flexible Compound Covariate Method (WFCCM) to verify if the selected proteins have statistically significant predictive power on the study subjects, and (3) employment of an agglomerative hierarchical clustering algorithm to investigate expression patterns among the statistically significant discriminator proteins as well as the biologic status.

Using the permutation t test, modified info score method (InfoScore), significance analysis of microarrays (SAM), weighted gene analysis, Fisher's exact test, and the Wilcoxon rank sum test, the most significant m/z species with different levels of expression among the

groups were selected. The cut off points for each method were $p < 0.0001$, 0, 3.5, 2, $p < 0.0001$ and $p < 0.0001$, respectively. Protein markers were included in the final list if they met at least 3 of these selection criteria. The misclassification rate was estimated using the leave-one-out cross-validated class prediction method. An agglomerative hierarchical clustering algorithm was employed to investigate patterns among the statistically significant discriminator proteins.

Protein identification

Protein extracts were prepared using 5 x 5 mm^3 portions of mammary tumor tissue in extraction buffer (0.25 M sucrose, 0.01 M Tris-HCl and 0.1 mM phenylmethanesulfonyl fluoride; Sigma), homogenized with a Dual homogenizer and centrifuged 3 times (10 min at 680 g, 10 min at 10,000 g and 1 hr at 100,000 g), each time transferring the soluble fraction to a new tube. The final fraction was separated over a C8 column (208TP54, Vydac, Hesperia, CA, USA) at 40°C using a Waters Alliance HPLC system (Milford, MA, USA). Solvent A was 0.1% trifluoroacetic acid (TFA) in water and solvent B was 0.085% TFA in acetonitrile. A flow rate of 1ml/min was used starting at 5% solvent B for 5 min, followed by a series of gradients: 5% to 25% B in 5 min, 25% to 60% B over 50 min, 60% to 95% B over 10 min. Collection ended with a final hold at 95% B for 10 min. The HPLC fractions were collected every minute with individual fractions monitored by MALDI-TOF MS. The fractions were then completely dried using Thermo Quest Savant Speedvac (Holbrook, NY, USA). Dried HPLC fractions were reconstituted in 10 µl of 5/5/0.3, v/v/v, acetonitrile/water/TFA and analyzed by MALDI-TOF MS. Fractions containing proteins of interest were digested with trypsin for identification. The samples were re-lyophilized and reconstituted with 10 microliters of 0.4 M ammonium hydrogen carbonate (EM Science). These fractions were reduced in incubation with 5 µl of 45 mM dithiothreitol (Sigma) at 60°C for 15 min, followed by alkylation with 5 µl of 100 mM iodoacetamide (Sigma) in the dark for 15 min. For protein digestion, the samples were trypsinized (1 µl of 1 µM bovine sequencing grade

trypsin, Promega) for 4 hrs at 37°C. The tryptic fragments were subjected to LC-MS/MS analysis using a ThermoFinnigan LTQ mass spectrometer (San Jose, CA, USA). Two microliter samples were loaded into a 100 μm i.d. self-packed micro-capillary reverse phase column packed with Monitor C18 -Spherical Silica from Column Engineering Inc (Ontario, CA, USA). The mobile phase A was 0.1% formic acid in water and phase B was 0.1% formic acid in acetonitrile. The gradient for mobile phase B started at 0% for 3 min, to 5% in 2 min, to 50% in 45 min, and to 90% in 5 min. The flow rate at the source was 700 nl /min. The tandem mass spectra were used to search the National Center for Biotechnology Information (NCBI) human protein database using the SEQUEST algorithm [27].

REFERENCES

1. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX*, et al.*: From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* 14:61-65, 1996

2. Washburn MP, Wolters D, Yates JR, 3rd: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology* 19:242-247, 2001

3. Beranova-Giorgianni S: Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. *TrAC, Trends in Analytical Chemistry* 22:273-281, 2003

4. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM*, et al.*: Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology* 17:676-682, 1999

5. Hochstrasser DF: *Proteome Research: New Frontiers in Functional Genomics*, Springer, Heidelberg, 1997

6. Eisenhaber F, Eisenhaber B, Maurer-Stroh S: Prediction of post-translational modifications from amino acid sequence: problems, pitfalls, and methodological hints. *Bioinformatics and Genomes*:81-105, 2003

7. Rowen L, Mahairas G, Hood L: Sequencing the human genome. *Science* 278:605-607, 1997

8. Fraser CM, Fleischmann RD: Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* 18:1207-1216, 1997

9. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM: Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64-71, 1989

10. Karas M, Bachmann D, Bahr U, Hillenkamp F: Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes* 78:53-68, 1987

11. Kriwacki RW, Siuzdak G: Probing protein-protein interactions with mass spectrometry. *Methods in Molecular Biology* 146:223-238, 2000

12. Akashi S: Studies on the protein structures by mass spectrometry. *Journal of the Mass Spectrometry Society of Japan* 45:1-23, 1997

13. Chaurand P, Schwartz SA, Caprioli RM: Profiling and imaging proteins in tissue sections by MS. *Analytical Chemistry* 76:86A-93A, 2004

14. Boguski MS: The turning point in genome research. *Trends in biochemical sciences* 20:295-296, 1995

15. Nair KS, Jaleel A, Asmann YW, Short KR, Raghavakaimal S: Proteomic research: Potential opportunities for clinical and physiological investigators. *American Journal of Physiology* 286:E863-E874, 2004

16. Mocellin S, Riccardo Rossi C, Traldi P, Nitti D, Lise M: Molecular oncology in the post-genomic era: the challenge of proteomics. *Trends in Molecular Medicine* 10:24-32, 2004

17. Zoon KC: Future Directions in Cancer Research: Impact of the Completion of the Human Genome. *Toxicologic Pathology* 32:1-2, 2004

18. Bukowska A, Lendeckel U, Kaehne T, Goette A: Proteomics in myocardial diseases. *Pathology, Research and Practice* 200:135-145, 2004

19. Butterfield DA: Proteomics: a new approach to investigate oxidative stress in Alzheimer's disease brain. *Brain Research* 1000:1-7, 2004

20. Ardekani AM, Liotta LA, Petricoin E, III: Clinical potential of proteomics in the diagnosis of ovarian cancer. *Expert Review of Molecular Diagnostics* 2:312-320, 2002

21. Petricoin EF, Liotta LA: Clinical applications of proteomics. *Journal of Nutrition* 133:2476S-2484S, 2003

22. Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR*, et al.*: Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 362:433-439, 2003

23. Schwartz SA, Weil RJ, Johnson MD, Toms SA, Caprioli RM: Protein profiling in brain tumors using mass spectrometry: Feasibility of a new technique for the analysis of protein expression. *Clinical Cancer Research* 10:981-987, 2004

24. Lilley KS, Razzaq A, Dupree P: Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Current Opinion in Chemical Biology* 6:46-50, 2002

25. Friedman DB, Hill S, Keller JW, Merchant NB, Levy SE, Coffey RJ, Caprioli RM: Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 4:793-811, 2004

26. Link AJ: Multidimensional peptide separations in proteomics. *Trends in Biotechnology* 20:S8-S13, 2002

27. Eng JK, McCormack AL, Yates JR, III: An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* 5:976-989, 1994

28. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* 17:994-999, 1999

29. Dunkley TPJ, Dupree P, Watson RB, Lilley KS: The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in Arabidopsis thaliana. *Biochemical Society Transactions* 32:520-523, 2004

30. Meehan KL, Sadar MD: Quantitative profiling of LNCaP prostate cancer cells using isotope-coded affinity tags and mass spectrometry. *Proteomics* 4:1116-1134, 2004

31. Chaurand P, Fouchecourt S, DaGue BB, Xu BJ, Reyzer ML, Orgebin-Crist MC, Caprioli RM: Profiling and imaging proteins in the mouse epididymis by imaging mass spectrometry. *Proteomics* 3:2221-2239, 2003

32. Chaurand P, Schwartz SA, Caprioli RM: Assessing protein patterns in disease using imaging mass spectrometry. *Journal of Proteome Research* 3:245-252, 2004

33. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM: Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nature medicine* 7:493-496, 2001

34. Chaurand P, Caprioli RM: Direct profiling and imaging of peptides and proteins from mammalian cells and tissue sections by mass spectrometry. *Electrophoresis* 23:3125-3135, 2002

35. Zhuang Z, Bertheau P, Emmert-Buck MR, Liotta LA, Gnarra J, Linehan WM, Lubensky IA: A microdissection technique for archival DNA analysis of specific cell populations in lesions < 1 mm in size. *American journal of pathology* 146:620-625., 1995

36. Meier-Ruge W, Bielser W, Remy E, Hillenkamp F, Nitsche R, Unsold R: The laser in the Lowry technique for microdissection of freeze-dried tissue slices. *Histochemical journal* 8:387-401, 1976

37. Shibata D, Hawes D, Li ZH, Hernandez AM, Spruck CH, Nichols PW: Specific genetic analysis of microscopic tissue after selective ultraviolet radiation fractionation and the polymerase chain reaction. *American journal of pathology* 141:539-543., 1992

38. Micke P, Bjornsen T, Scheidl S, Stromberg S, Demoulin J-B, Ponten F, Ostman A*, et al.*: A fluid cover medium provides superior morphology and preserves RNA integrity in tissue sections for laser microdissection and pressure catapulting. *Journal of pathology* 202:130-138., 2004

39. Irie T, Aida T, Tachikawa T: Gene expression profiling of oral squamous cell carcinoma using laser microdissection and cDNA microarray. *Medical Electron Microscopy* 37:89-96, 2004

40. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhang Z, Goldstein SR, Weiss RA*, et al.*: Laser capture microdissection. *Science* 274:998-1001, 1996

41. Rubin MA: Use of laser capture microdissection, cDNA microarrays, and tissue microarrays in advancing our understanding of prostate cancer. *Journal of Pathology* 195:80-86, 2001

42. Suzuki K, Matsui H, Yamanaka H: Laser capture microdissection for microscopic analysis and detection of RNA and DNA for genetic analysis. *Soshiki Saibo Kagaku Koshukai* 27th:127-135, 2002

43. Hoang CD, D'Cunha J, Tawfic SH, Gruessner AC, Kratzke RA, Maddaus MA: Expression profiling of non-small cell lung carcinoma identifies metastatic genotypes based on lymph node tumor burden. *Journal of Thoracic and Cardiovascular Surgery* 127:1332-1342, 2004

44. Coon KD, Dunckley T, Stephan DA: Biomarker identification in neurologic diseases: improving diagnostics and therapeutics. *Expert Review of Molecular Diagnostics* 4:361-375, 2004

45. Schmid H, Henger A, Cohen CD, Frach K, Groene H-J, Schloendorff D, Kretzler M: Gene expression profiles of podocyte-associated molecules as diagnostic markers in acquired proteinuric diseases. *Journal of the American Society of Nephrology* 14:2958-2966, 2003

46. Torres-Munoz J, Stockton P, Tacoronte N, Roberts B, Maronpot RR, Petito CK: Detection of HIV-1 gene sequences in hippocampal neurons isolated from postmortem AIDS brains by laser capture microdissection. *Journal of neuropathology and experimental neurology* 60:885-892, 2001

47. Hong SH, Nah HY, Lee JY, Gye MC, Kim CH, Kim MK: Analysis of estrogen-regulated genes in mouse uterus using cDNA microarray and laser capture microdissection. *Journal of Endocrinology* 181:157-167, 2004

48. Emmert-Buck MR, Gillespie JW, Paweletz CP, Ornstein DK, Basrur V, Appella E, Wang QH*, et al.*: An approach to proteomic analysis of human tumors. *Molecular carcinogenesis* 27:158-165., 2000

49. Wulfkuhle Julia D, Sgroi Dennis C, Krutzsch H, McLean K, McGarvey K, Knowlton M, Chen S*, et al.*: Proteomics of human breast ductal carcinoma in situ. *Cancer research* 62:6740-6749., 2002

50. Palmer-Toy DE, Sarracino DA, Sgroi D, LeVangie R, Leopold PE: Direct acquisition of matrix-assisted laser desorption/ionization time-of-flight mass spectra from laser capture microdissected tissues. *Clinical Chemistry* 46:1513-1516, 2000

51. Xu BJ, Caprioli RM, Sanders ME, Jensen RA: Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *Journal of the American Society for Mass Spectrometry* 13:1292-1297, 2002

52. Chaurand P, Schwartz SA, Billheimer D, Xu BJ, Crecelius A, Caprioli RM: Integrating histology and imaging mass spectrometry. *Analytical Chemistry* 76:1145-1155, 2004

53. Jemal A, Tiwari Ram C, Murray T, Ghafoor A, Samuels A, Ward E, Feuer Eric J*, et al.*: Cancer statistics, 2004. *CA: a cancer journal for clinicians* 54:8-29, 2004

54. Cai WB, Roberts SA, Potten CS: The number of clonogenic cells in crypts in three regions of murine large intestine. *International Journal of Radiation Biology* 71:573-579, 1997

55. Potten CS: Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 353:821-830, 1998

56. Marshman E, Booth C, Potten CS: The intestinal epithelial stem cell. *BioEssays : news and reviews in molecular, cellular and developmental biology* 24:91-98, 2002

57. Booth C, Potten CS: Gut instincts: thoughts on intestinal epithelial stem cells. *Journal of Clinical Investigation* 105:1493-1499, 2000

58. Moser AR, Pitot HC, Dove WF: A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* 247:322-324, 1990

59. Kinzler KW, Vogelstein B: Lessons from hereditary colorectal cancer. *Cell* 87:159-170, 1996

60. Seidensticker MJ, Behrens J: Biochemical interactions in the wnt pathway. *Biochimica et biophysica acta* 1495:168-182, 2000

61. Huber O, Korn R, McLaughlin J, Ohsugi M, Herrmann BG, Kemler R: Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. *Mechanisms of development* 59:3-10, 1996

62. Munemitsu S, Souza B, Mueller O, Albert I, Rubinfeld B, Polakis P: The APC gene product associates with microtubules in vivo and promotes their assembly in vitro. *Cancer Research* 54:3676-3681, 1994

63. Boivin Gregory P, Washington K, Yang K, Ward Jerrold M, Pretlow Theresa P, Russell R, Besselsen David G*, et al.*: Pathology of mouse models of intestinal cancer: consensus report and recommendations. *Gastroenterology* 124:762-777, 2003

64. Shyr Y, Kim K: Weighted flexible compound covariate method for classifying microarray data: a case study of gene expression level of lung cancer. *Practical Approach to Microarray Data Analysis*:186-200, 2003

65. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* 95:14863-14868, 1998

66. Komatsu K, Andoh A, Ishiguro S, Suzuki N, Hunai H, Kobune-Fujiwara Y, Kameyama M*, et al.*: Increased expression of S100A6 (calcyclin), a calcium-binding protein of the S100 family, in human colorectal adenocarcinomas. *Clinical Cancer Research* 6:172-177, 2000

67. Komatsu K, Murata K, Kameyama M, Ayaki M, Mukai M, Ishiguro S, Miyoshi J*, et al.*: Expression of S100A6 and S100A4 in matched samples of human colorectal mucosa, primary colorectal adenocarcinomas and liver metastases. *Oncology* 63:192-200, 2002

68. Lewington AJ, Padanilam BJ, Hammerman MR: Induction of calcyclin after ischemic injury to rat kidney. *American journal of physiology* 273:F380-385, 1997

69. Tonini GP, Casalaro A, Cara A, Di Martino D: Inducible expression of calcyclin, a gene with strong homology to S-100 protein, during neuroblastoma cell differentiation and its prevalent expression in Schwann-like cell lines. *Cancer research* 51:1733-1737, 1991

70. Weterman MA, van Muijen GN, Bloemers HP, Ruiter DJ: Expression of calcyclin in human melanocytic lesions. *Cancer research* 53:6061-6066:, 1993

71. Weterman MA, Stoopen GM, van Muijen GN, Kuznicki J, Ruiter DJ, Bloemers HP: Expression of calcyclin in human melanoma cell lines correlates with metastatic behavior in nude mice. *Cancer research* 52:1291-1296, 1992

72. Courtois-Coutry N, Le Moellic C, Boulkroun S, Fay M, Cluzeaud F, Escoubet B, Farman N*, et al.*: Calcyclin is an early vasopressin-induced gene in the renal collecting duct. Role in the long term regulation of ion transport. *Journal of Biological Chemistry* 277:25728-25734, 2002

73. Ebralidze A, Tulchinskii E, Grigoryan M, Afanas'eva A, Senin V, Revazova E, Lukanidin E: Isolation and characterization of a gene specifically expressed in different metastatic cells and whose deduced gene product has a high degree of homology to a calcium-binding protein family. *Genes & Development* 3:1086-1093, 1989

74. Pedrocchi M, Schafer BW, Mueller H, Eppenberger U, Heizmann CW: Expression of Ca(2+)-binding proteins of the S100 family in malignant human breast-cancer cell lines and biopsy samples. *International journal of cancer* 57:684-690:, 1994

75. Chaurand P, DaGue BB, Pearsall RS, Threadgill DW, Caprioli RM: Profiling proteins from azoxymethane-induced colon tumors at the molecular level by matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* 1:1320-1326, 2001

76. Stulik J, Osterreicher J, Koupilova K, Knizek, Macela A, Bures J, Jandik P*, et al.*: The analysis of S100A9 and S100A8 expression in matched sets of macroscopically normal colon mucosa and colorectal carcinoma: the S100A9 and S100A8 positive cells underlie and invade tumor mass. *Electrophoresis* 20:1047-1054., 1999

77. Poullis A, Foster R, Mendall MA, Fagerhol MK: Emerging role of calprotectin in gastroenterology. *Journal of Gastroenterology and Hepatology* 18:756-762, 2003

78. Isaksen B, Fagerhol MK: Calprotectin inhibits matrix metalloproteinases by sequestration of zinc. *Molecular Pathology* 54:289-292, 2001

79. Johne B, Fagerhol MK, Lyberg T, Prydz H, Brandtzaeg P, Naess-Andresen CF, Dale I: Functional and clinical aspects of the myelomonocyte protein calprotectin. *Molecular Pathology* 50:113-123, 1997

80. Roseth AG, Kristinsson J, Fagerhol MK, Schjonsby H, Aadland E, Nygaard K, Roald B: Faecal calprotectin: a novel test for the diagnosis of colorectal cancer? *Scandinavian journal of gastroenterology* 28:1073-1076, 1993

81. Roseth AG: Determination of faecal calprotectin, a novel marker of organic gastrointestinal disorders. *Digestive and liver disease* 35:607-609, 2003

82. Lahm H, Andre S, Hoeflich A, Kaltner H, Siebert H-C, Sordat B, von der Lieth C-W*, et al.*: Tumor galectinology: Insights into the complex network of a family of endogenous lectins. *Glycoconjugate Journal* 20:227-238, 2004

83. Hittelet A, Legendre H, Nagy N, Bronckart Y, Pector J-C, Salmon I, Yeaton P*, et al.*: Upregulation of galectins-1 and -3 in human colon cancer and their role in regulating cell migration. *International Journal of Cancer* 103:370-379, 2003

84. Schaffert C, Pour PM, Chaney WG: Localization of galectin-3 in normal and diseased pancreatic tissue. *International journal of pancreatology* 23:1-9, 1998

85. Gillenwater A, Xu XC, el-Naggar AK, Clayman GL, Lotan R: Expression of galectins in head and neck squamous cell carcinoma. *Head & neck* 18:422-432., 1996

86. Schoeppner HL, Raz A, Ho SB, Bresalier RS: Expression of an endogenous galactose-binding lectin correlates with neoplastic progression in the colon. *Cancer* 75:2818-2826, 1995

87. Nakamura M, Inufusa H, Adachi T, Aga M, Kurimoto M, Nakatani Y, Wakano T*, et al.*: Involvement of galectin-3 expression in colorectal cancer progression and metastasis. *International journal of oncology* 15:143-148., 1999

88. Whitehead RH, Brown A, Bhathal PS: A method for the isolation and culture of human colonic crypts in collagen gels. *In Vitro Cell Dev Biol.* 23:436-442, 1987

89. Fogo AB, Hawkins EP, Berry PL, Glick AD, Chiang ML, MacDonell RC, Jr., Ichikawa I: Glomerular hypertrophy in minimal change disease predicts subsequent progression to focal glomerular sclerosis. *Kidney international* 38:115-123, 1990

90. Fogo AB: Glomerular hypertension, abnormal glomerular growth, and progression of renal diseases. *Kidney international* 75:S15-21, 2000

91. Klein JB, Thongboonkerd V: Overview of proteomics. *Contributions to Nephrology* 141:1-10, 2004

92. Arthur JM: Proteomics. *Current Opinion in Nephrology and Hypertension* 12:423-430, 2003

93. Ma L-J, Nakamura S, Whitsitt JS, Marcantoni C, Davidson JM, Fogo AB: Regression of sclerosis in aging by an angiotensin inhibition-induced decrease in PAI-1. *Kidney International* 58:2425-2436, 2000

94. Kerins DM, Hao Q, Vaughan DE: Angiotensin induction of PAI-1 expression in endothelial cells is mediated by the hexapeptide angiotensin IV. *Journal of clinical investigation* 96:2515-2520, 1995

95. Good KS, O'Brien K, Schulman G, Kerjaschki D, Fogo AB: Unexplained nephrotic-range proteinuria in a 38-year-old man: a case of "no change disease". *American journal of kidney diseases* 43:933-938, 2004

96. Safer D, Golla R, Nachmias VT: Isolation of a 5-kilodalton actin-sequestering peptide from human blood platelets. *Proceedings of the National Academy of Sciences of the United States of America* 87:2536-2540, 1990

97. Grant DS, Rose W, Yaen C, Goldstein A, Martinez J, Kleinman H: Thymosin beta4 enhances endothelial cell differentiation and angiogenesis. *Angiogenesis* 3:125-135, 1999

98. Al-Nedawi KNI, Czyz M, Bednarek R, Szemraj J, Swiatkowska M, Cierniewska-Cieslak A, Wyczolkowska J*, et al.*: Thymosin b4 induces the synthesis of plasminogen activator inhibitor 1 in cultured endothelial cells and increases its extracellular expression. *Blood* 103:1319-1324, 2004

99. Eddy AA: Plasminogen activator inhibitor-1 and the kidney. *American Journal of Physiology* 283:F209-220, 2002

100. Ma L-j, Fogo AB: Role of angiotensin II in glomerular injury. *Seminars in Nephrology* 21:544-553, 2001

101. Ikoma M, Kawamura T, Kakinuma Y, Fogo AB, Ichikawa I: Cause of variable therapeutic efficiency of angiotensin converting enzyme inhibitor on glomerular lesions. *Kidney International* 40:195-202, 1991

102. Mora CA, Baumann CA, Paino JE, Goldstein AL, Badamchian M: Biodistribution of synthetic thymosin beta 4 in the serum, urine, and major organs of mice. *International journal of immunopharmacology* 19:1-8, 1997

103. Akis N, Madaio MP: Isolation, culture, and characterization of endothelial cells from mouse glomeruli. *Kidney International* 65:2223-2227, 2004

104. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98:5116-5121., 2001

105. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P*, et al.*: Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 344:539-548., 2001

106. Kaminiski N, Friedman N: Practical approaches to analyzing results of microarray experiments. *American Journal of Respiratory Cell and Molecular Biology* 27:125-132, 2002

107. Ben-Dor A, Friedman N, Yakhini Z: Scoring genes for relevance, in, Tech Report AGL-2000-13, Agilent Labs, Agilent Technologies, 2000, http://www.labs.agilent.com/resources/techreports.html.

108. Tukey JW: Tightening the clinical trial. *Control Clin Trials* 14:266-285., 1993

109. Shyr Y, KyungMann K: Weighted Flexible Compound Covariate Method for Classifying Microarray Data, in *A Practical Approach to Microarray Data Analysis*, edited by Berrar D, Norwell, MA, Kluwer Academic Publishers, 2003

110. Yamagata N, Shyr Y, Yanagisawa K, Edgerton M, Dang TP, Gonzalez A, Nadaf S*, et al.*: A Training-Testing Approach to the Molecular Classification of Resected Non-Small Cell Lung Cancer. *Clinical Cancer Research* 9:4695-4704, 2003

111. Allred DC, Mohsin SK, Fuqua SA: Histological and biological evolution of human premalignant breast disease. *Endocrine-related cancer* 8:47-61, 2001

112. Ingvarsson S: Molecular genetics of breast cancer progression. *Seminars in cancer biology* 9:277-288, 1999

113. Otto AM, Mueller CSG, Huff T, Hannappel E: Chemotherapeutic drugs change actin skeleton organization and the expression of b-thymosins in human breast cancer cells. *Journal of Cancer Research and Clinical Oncology* 128:247-256, 2002

114. Yamamoto T, Gotoh M, Kitajima M, Hirohashi S: Thymosin beta-4 expression is correlated with metastatic capacity of colorectal carcinomas. *Biochemical and biophysical research communications* 193:706-710, 1993

115. Goldstein AL: Thymosin b4: A new molecular target for antitumor strategies. *Journal of the National Cancer Institute* 95:1646-1647, 2003

116. Cha H-J, Jeong M-J, Kleinman HK: Role of Thymosin b4 in Tumor Metastasis and Angiogenesis. *Journal of the National Cancer Institute* 95:1674-1680, 2003

117. Schmitt M, Harbeck N, Thomssen C, Wilhelm O, Magdolen V, Reuning U, Ulm K*, et al.*: Clinical impact of the plasminogen activation system in tumor invasion and metastasis. Prognostic relevance and target for therapy. *Thrombosis and Haemostasis* 78:285-296, 1997

118. Andreasen PA, Kjoller L, Christensen L, Duffy MJ: The urokinase-type plasminogen activator system in cancer metastasis: a review. *International journal of cancer* 72:1-22, 1997

119. Song JJ, Rhee JG, Suntharalingam M, Walsh SA, Spitz DR, Lee YJ: Role of Glutaredoxin in Metabolic Oxidative Stress. Glutaredoxin as a Sensor of Oxidative Stress Mediated by H2O2. *Journal of Biological Chemistry* 277:46566-46575, 2002

120. Daily D, Vlamis-Gardikas A, Offen D, Mittelman L, Melamed E, Holmgren A, Barzilai A: Glutaredoxin protects cerebellar granule neurons from dopamine-induced apoptosis by activating NF-kB via Ref-1. *Journal of Biological Chemistry* 276:1335-1344, 2001

121. Jordan A, Reichard P: Ribonucleotide reductases. *Annual Review of Biochemistry* 67:71-98, 1998

122. Hirota K, Matsui M, Murata M, Takashima Y, Cheng FS, Itoh T, Fukuda K*, et al.*: Nucleoredoxin, glutaredoxin, and thioredoxin differentially regulate NF-kappaB, AP-1, and CREB activation in HEK293 cells. *Biochemical and biophysical research communications* 274:177-182, 2000

123. Nardini M, Mazzocco M, Massaro A, Maffei M, Vergano A, Donadini A, Scartezzini P*, et al.*: Crystal structure of the glutaredoxin-like protein SH3BGRL3 at 1.6 Angstrom resolution. *Biochemical and biophysical research communications* 318:470-476, 2004

124. Deng G, Lu Y, Zlotnikov G, Thor AD, Smith HS: Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* 274:2057-2059, 1996