HIGH-THROUGHPUT ANALYSIS OF TISSUE MICROARRAYS OF DISEASE: COMBINING *IN SITU* PROTEOMICS WITH MALDI IMAGING MASS SPECTROMETRY

By

Mark Reid Groseclose

Dissertation Submitted to the Faculty of the Graduate School of Vanderbilt
University in fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
In
Chemistry
August, 2009

Approved:

Professor Richard M. Caprioli

Professor Darryl J. Bornhop

Professor David L. Hachey

Professor Eva M. Harth

This work is dedicated to my father and mother, Mark and Gwyn Groseclose.
Thank you for everything.

ACKNOWLEDGEMENTS

Over the past five years at Vanderbilt University, I have met and worked with a number of truly remarkable people. In retrospect, I realize that I had no idea what I was getting myself into when I began the Ph.D program. Through the highs and lows, I can now say that it has been a very rewarding experience and there are several people that I would like to acknowledge.

First, I would like to thank my advisor, Richard Caprioli, for his support, guidance, and motivation as a mentor. The MSRC is a unique establishment with an exceptional number of resources and is undoubtedly one of the most advanced labs in the field of mass spectrometry. I am extremely grateful for being given an opportunity to be a part of this center and my time here has provided me with a solid foundation and background on which to continue my career. I would also like to thank the remaining members of my Ph.D committee, Darryl Bornhop, David Hachey, and Eva Harth, for their time commitment and support.

Next, I would like to acknowledge those that helped to progress my research and directly impacted the success of this project. First and foremost, I would like to thank Pierre Massion for providing the FFPE lung tissue samples and for his support and guidance throughout this work. Having insight from a clinical perspective helped to define the direction that this project took from early on and shed light on the impact that it may someday have in the medical field.

Finally, on a personal note, thank you to my family, including my parents, Mark and Gwyn, my brother and sister, Chase and Brooke, and my grandparents, Robert and June Groseclose and Charles and Betty Smallwood, and my wonderful girlfriend, Audrey. Your love and support through the years has always been a source of inspiration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TECHNOLOGICAL ABBREVIATIONS

MS, Mass Spectrometry

IMS, Imaging Mass Spectrometry

MALDI, Matrix Assisted Laser Desorption/Ionization

ESI, Electrospray Ionization

FTICR, Fourier Transform Ion Cyclotron Resonance

TOF, Time-Of-Flight

MS/MS, Tandem-MS

m/z, Mass-to-Charge Ratio

Da, Dalton

TIC, Total Ion Current

LC, Liquid Chromatography

SA, Sinapinic Acid

DHB, 2, 5-Dihydroxybenzoic Acid

CHCA, α-Cyano-4-Hydroxycinnamic Acid

IHC, Immunohistochemistry

FFPE, Formalin-Fixed Paraffin-Embedded

NSCLC, Non-Small Cell Lung Cancer

TMA, Tissue Microarray

LIFT Device, refers to the four grid system in the MALDI-TOF/TOF instrument that serves to focus and reaccelerate ion packets for tandem MS

CHAPTER I

INTRODUCTION

Biomarker Discovery

The NIH defined biomarkers as characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic interventions.[1] In a disease such as cancer the identification of a molecule or molecular signature that is accurately indicative of these processes will be of extraordinary benefit to clinicians and patients.[2]

The main hypothesis driving the search for cancer biomarkers is the concept that organs secrete a specific set of proteins representing a molecular signature indicative of its physiological state.[3] In cancer patients this set of secreted proteins should alter to reflect the genetic mutations and other processes that contribute to the cancer phenotype.[3] Detection and characterization of these molecular fingerprints has begun to provide a unique view of the proteomic changes associated with disease status. The potential use of biomarkers for the early detection of cancer has compelled significant research in this field.[2, 4]

Another advantage of biomarkers is the potential to aid clinicians in selecting patients to undergo certain treatments. This research is conducted based on previous evaluations of treatment efficacy and safety from patients

exhibiting a specific biomarker. These markers may also be used to monitor response to treatment and disease progression.[5]

Various biological specimens, including blood, urine, and saliva, have been analyzed for the discovery of potential biomarker candidates. One of the major goals in this field of research is the development of a blood-based assay for the detection of biomarkers, due to the relatively easy and non-invasive manner in which blood can be collected. Blood, however, poses many significant challenges to researchers during the discovery phase. The range of protein concentrations in blood is extensive, ranging from albumin at ~40 mg/mL, to cytokines at ~5 pg/mL.[2] Moreover, a set of 22 proteins account for 99% of the total protein content in blood.[6, 7] This compositional complexity makes the detection of low abundance proteins extremely difficult when using common separation techniques such as 2-DIGE and HPLC.

Immunodepletion technologies were developed to reduce the presence of the high abundance proteins in complex samples.[6] Although these depletion strategies have proven effective, they are not without limitations. For example, immunodepletion of albumin from serum with a removal efficiency of 99% leaves a remaining concentration of approximately 50 μg/mL, which is still significantly higher than most proteins of interest. Furthermore, non-specific removal of proteins other than the target and possible binding of target proteins to other proteins, like albumin, will inevitably result in concomitant depletion of potentially useful markers. The extremely dynamic nature of the plasma proteome adds yet another variable to consider when profiling these samples. Factors such as

genetic polymorphisms, gender, age, ethnicity, and lifestyle can significantly alter the protein species detected in blood. Variations arising from inconsistent sample collection and storage procedures (e.g. processing time, storage temperature) may further exacerbate these fluctuations in the proteome.

These shortcomings have encouraged research looking at alternative sources, such as tumor biopsy tissues, for biomarker discovery. In theory, the concentrations of potential biomarkers are expected to be highest in the tumor and the microenvironment of the surrounding tissue. Upon entry into circulation via the lymphatic system, these markers will be significantly diluted into the complex matrix of the blood. Therefore, the direct analysis of the tumor and surrounding tissue should be the most logical source to carry out the discovery phase of biomarkers. The markers identified in this way could then be detected in blood using highly specific technologies such as multiple reaction monitoring (MRM) or immunoassays.

Imaging Mass Spectrometry

Molecular imaging has played a pivotal role in our understanding of the spatial complexity of biological systems at the tissue and cellular level.[8] Several different approaches are commonly used to examine the spatial distribution of molecules in biological samples including fluorescence[9], immunohistochemistry (IHC)[10], positron emission topography[11], and magnetic resonance imaging (MRI)[12]. Although these technologies represent powerful and well-established imaging tools, there are several limitations present with these approaches. Most

notably, these methodologies generally employ a targeted approach that requires an *a priori* knowledge of the molecule(s) of interest, thus limiting their effectiveness as true discovery tools. Furthermore, these techniques are only capable of analyzing a small number of components simultaneously and therefore provide a limited, albeit valuable, view of the biological system.

Over the past several years, imaging mass spectrometry (IMS) has emerged as a powerful tool for studying the spatial arrangement of proteins, peptides, lipids, and small molecules in tissues.[13-15] The multichannel detection capability of mass spectrometry enables the position sensitive analysis of hundreds of different molecules in a single experiment.[13, 14] This is achieved by acquiring mass spectra across a sample at precisely defined geometrical coordinates. Post acquisition processing compiles the mass spectra into a format in which any of the detected species can be viewed as an ion density map, where the relative abundance of the selected ion across the sample is displayed on a color intensity scale at each coordinate location (pixel) (**Figure 1**). Unlike other molecular visualization techniques, IMS does not require a target specific reagent and it is therefore a valuable discovery tool.

History

IMS has been conducted on a wide range of analytes using a variety of different ionization and mass analyzer combinations. As early as the 1960s, secondary ion mass spectrometry (SIMS) was used to examine the distribution of
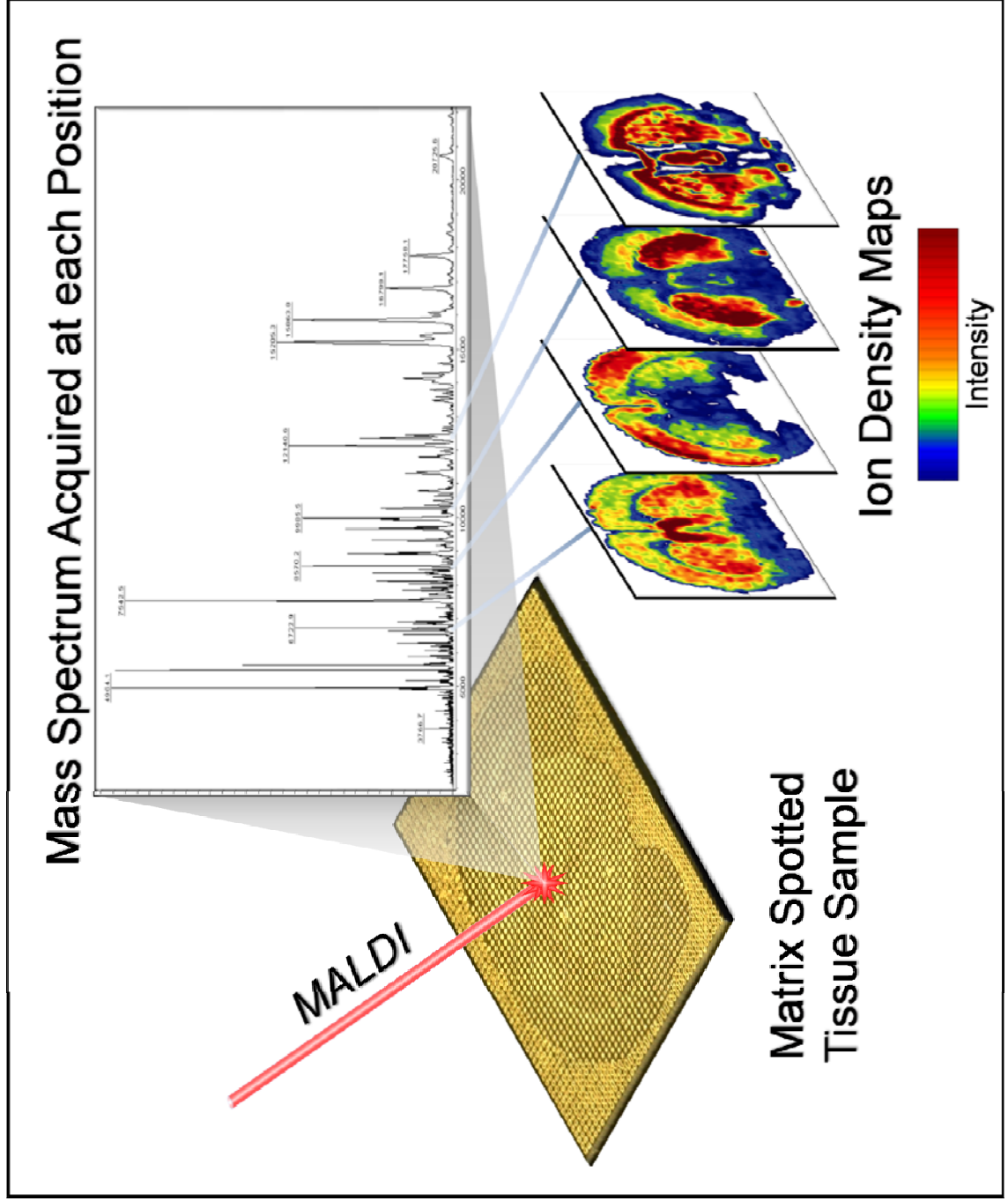
Figure 1. Visual representation of the workflow for a MALDI Imaging MS experiment.

light elements (Na^+, K^+, Ca^{2+}) or exogenous, isotopically labeled elements.[16] The desorption and ionization process in SIMS is a highly energetic process that results in extensive molecular fragmentation, thereby limiting the practical mass range to small molecule analysis. During the 1990s, polyatomic primary ion beams and the liquid metal ion gun were developed to extend the effective mass range of SIMS by providing a softer ionization process, which allowed analysis of cholesterol, lipids, and other small molecules up to ~1,000 Da.[17-19] Despite significant advances in the ion generation process in SIMS instruments, larger biological molecules such as peptides, proteins, and oligonucleotides cannot be effectively analyzed using this technique. It should be noted, however, that for small molecule imaging experiments SIMS provides superior spatial resolution relative to alternative techniques and is capable of imaging sub-cellular components.[20]

With the introduction of matrix-assisted laser desorption/ionization (MALDI) in the mid 1980s [21, 22] intact high molecular weight analytes could be efficiently desorbed and ionized for detection in a mass spectrometer. Along with electrospray ionization[23] (ESI), these soft ionization techniques would become the basis for a revolution in bioanalytical mass spectrometry and advance the field into a new era.

In 1997, Caprioli et al.[24] demonstrated that MALDI MS could be used to map the location and relative intensity of proteins and peptides in a tissue section. In the years since this pioneering work, IMS has evolved into a highly versatile technique with a broad range of functionality and applications.[14, 25, 26]

The Technology

A brief description of the general workflow for an imaging experiment will be described in the following text; more detailed information can be found in several previously published works. [27, 28] It should be noted that a complimentary technique termed direct tissue profiling, is carried out in exactly the same manner as an imaging experiment with the exception that the data is not necessarily acquired in an arrayed pattern for generating ion density maps. Profiling experiments are carried out when the desired area of analysis is small, and mass spectra acquisition is only necessary at a limited number of positions.

Tissue Collection

To begin, a fresh frozen or formalin-fixed paraffin-embedded (FFPE) tissue specimen (e.g. tumor biopsy) is cut into thin sections (~12 μm for frozen; ~5 μm for FFPE) and mounted onto a conductive MALDI target, typically a gold coated plate or indium tin oxide (ITO) coated microscope slide. Frozen tissues are sectioned in a cryostat and are thaw mounted onto the target. The tissues and MALDI target are slowly equilibrated to room temperature and then dried in a vacuum desiccator. For protein and peptide IMS experiments, these adhered sections are typically washed using a series of graded ethanol solutions to remove salts, lipids, and other contaminants that may result in ion suppression.[29] Some IMS experiments, such as those specifically targeting lipids, do not employ these wash steps because they may remove the molecules of interest.[30]

FFPE tissues are sectioned at room temperature using a microtome. The sections are floated in a temperature controlled water bath (37° C) and then mounted onto the MALDI target. The FFPE sections are then deparaffinized and antigen retrieved prior to MS analysis (these steps will be described in more detail in Chapter 3).

Matrix Application

The most commonly used MALDI matrices include 3,5-dimethoxy-4-hydroxy-cinnamic acid (sinnapinic acid, SA), 2,5-dihydroxybenzoic acid (DHB), and α-cyano-4-hydroxycinnamic acid (CHCA). The matrix and solvent composition used will vary depending on the specific molecule and/or tissue being investigated. For example, SA is the most suitable matrix to be used in protein imaging experiments (m/z > 2000); whereas, DHB and CHCA provide superior performance in the low molecular weight mass range of peptides, lipids, and small molecules (m/z 100-2000). These findings have been mostly empirical and the fundamental mechanism behind the performance of the different matrices is not well understood, yet more information regarding this topic can be found in several recently published articles.[31-34]

Matrix application can be carried out using several different approaches. The two most commonly used approaches are spray coating, which applies a homogenous layer of matrix across the tissue, or spotting, which generates an array of discrete matrix spots across the sample. The advantage of spray coating is that the matrix is applied as a homogenous coating across the sample

allowing mass spectrometric acquisition to be carried out at high spatial resolutions limited only by the spot size of the laser. Spray coated matrix applications, however, often result in poor signal quality due to insufficient analyte extraction from the tissue during the relatively short solvent drying time. Furthermore, replication of exact spray conditions (e.g. humidity) and volumes, especially when carried out manually, can be difficult to achieve resulting in poor reproducibility between samples.

Alternatively, microspotting can be used to apply discrete droplets of matrix into defined arrays across a sample. Microspotting enables precise volumes to be reproducibly deposited at each position over several iterations. This is important when comparing the relative ion intensities across a sample, because inconsistencies in the matrix application will lead to artificial variation in signal intensities.

Two commercially available robotic spotters were used in this research. The Portrait 630 (Labcyte; Sunnyvale, CA), shown in **Figure 2a**, uses acoustic droplet ejection technology to emit a sound wave at a precisely defined power and focus into a solution reservoir, ejecting a ~170 pL droplet onto the adjacent sample surface. This instrument is nozzle-free and droplets can be ejected at a frequency of up to 200 Hz. The Portrait offers two separate working modes: start/stop mode, which can be used to deposit multiple drops at each position, and flyby mode, which moves the sample stage in a continuous motion while depositing a single drop at each position. When printing in flyby mode, a single
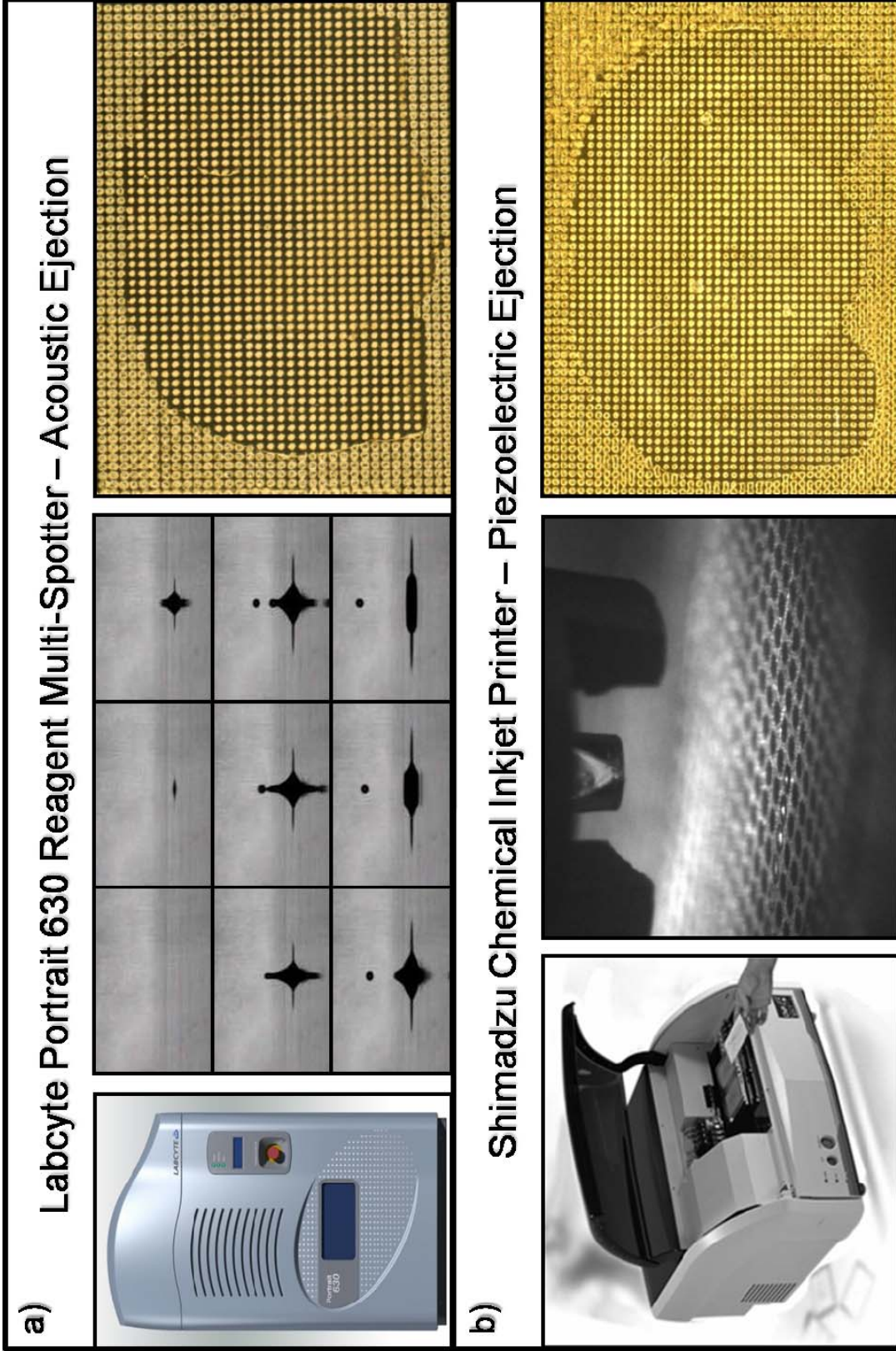
Figure 2. a) Labcyte Portrait 630 – instrument image and drop formation images adapted from http://www.labcyte.com/Portrait®_630_Spotter/Default.92.html **b.)** Shimadzu chemical inkjet printer (ChIP) – instrument image modified from http://www.shimadzu-biotech-.net/pages/products/2/chip.php

droplet ejection pattern covering approximately 1 cm^2 can be completed in less than 2 min.

The other instrument is a piezoelectric based chemical inkjet printer (ChIP; Shimadzu; Kyoto, Japan), shown in **Figure 2b**, that ejects ~80 pL droplets from a print head nozzle at up to 200 Hz. The ChIP only operates in start/stop mode and, as a result, printing can take significantly longer compared to the Portrait. For example, printing a grid pattern of spots with 250 μm center-to-center spacing covering 1 cm^2 would take approximately 6 min. with the ChIP compared to 2 min. with the Portrait. In flyby mode the printing time for a given area using the Portrait is constant, regardless of the spot to spot distance, because the stage motion is constant and the drop ejection frequency is varied. Because the ChIP only operates in start/stop mode, the printing time for a given area using this instrument will increase as the spot-to-spot distance decreases. Therefore, the major advantage of the Portrait is speed and the ability to use high matrix concentrations. Although the nozzle-free setup circumvents the risk of clogging, it allows the droplets to follow a slightly more variable trajectory onto the sample target resulting in slightly decreased accuracy and precision. The major advantage of the ChIP is that the nozzle based ejection provides a well-defined droplet trajectory onto the sample resulting in a high precision and high accuracy placement Also, the drop size on the ChIP is around half that of the Portrait, which enables arrays to be printed with closer spacing without droplet pooling. However, the piezoelectric droplet ejection parameters on the ChIP, such as voltage and pulse timing, can vary greatly between different solutions and tuning

can be very tedious. Furthermore, the nozzle can easily become clogged due to matrix crystallization and/or contaminants present in the matrix solution resulting in erratic drop ejection

In the experiments reported in this work, the ChIP was used for developing the *in situ* digestion method and the Portrait was used for applying this method to the lung cancer FFPE-TMA samples. As discussed in further detail in the following chapters, the robotic spotters are integral to this work. This is because when printing enzyme and matrix solutions, it is extremely important that precisely the same volume of solution is deposited at each position to ensure the resulting protein digestion is accurately representative of the protein content at each position.

Instrumentation

The most commonly used IMS platform is the MALDI-TOF mass spectrometer. Several other types of instruments including ion trap, FT-ICR, q-TOF, ion mobility, and DESI mass spectrometers have been used to perform imaging experiments and more information on these can be found in several recent publications.[26, 35, 36] The research presented in this work was performed using a MALDI-TOF and MALDI-TOF/TOF MS and therefore these will be the focus of the introduction.

MALDI, as it is currently used, was first reported by Karas and Hillenkamp in 1987.[37] In this seminal work, the matrix effect for enhancing the laser desorption ionization (LDI) of nonvolatile compounds was demonstrated using an

organic acid (nicotinic acid) that exhibited strong resonance absorption at the wavelength of the laser used. The ability of MALDI to produce high mass ions for mass spectrometry has led to its application in a broad range of biomolecular analyses including proteomics[38], oligonucleotides[39, 40], polysaccharides[41], glycoproteins[42-44], and synthetic polymers[45, 46].

Despite the widespread use of this technology, the exact role of the matrix is not fully understood. However, three key functions of the matrix were proposed in early papers[47]: 1) incorporation of the analyte molecules into the matrix crystals; 2) a collective laser absorption and desorption event; 3) analyte ionization through matrix interaction. Current models describing the UV-MALDI ionization mechanism are based on a 2-step framework involving an initial (primary) ion formation followed by ion-molecule reactions in the desorption/ablation plume that give rise to secondary ions. Several reviews are available which discuss these mechanisms in detail.[47-50]

The MALDI-TOF ionization/mass analyzer combination is well-suited for IMS for several reasons. Most importantly, the duty cycle of the TOF mass spectrometer is ideal to be coupled to the pulsed laser process of MALDI. MALDI, which uses laser ablation to carry out desorption/ionization of analytes, can be carried out at high repetition rates (up 5 kHz) due to the short pulse widths of solid state lasers (e.g. Nd:YAG) and the availability of high speed instrument electronics. This pulsed configuration provides enhanced sensitivity as compared to a scanning mass analyzer, such as an ion trap, because all ions are detected virtually simultaneously in each laser pulse. Other reasons

including a wide detectable mass range, high ion transmission efficiency, and simplicity in instrument design and maintenance make MALDI-TOF instruments ideal for most imaging experiments.

In MALDI-TOF instruments, ions formed following each laser pulse are accelerated by an electric field to the same kinetic energy into a field free drift region. Ions of different m/z values are separated according to differences in their resulting velocity, based on the equation:

$$T = k \cdot \sqrt{\frac{m}{q}}$$

where T is time-of-flight, m is mass, q is charge, and k is proportionality constant related to instrument dimensions and operating potentials. Therefore, for two ions with the same charge but different masses, the ion with the smaller mass will travel at a higher velocity than the one with the larger mass resulting in a separation in space and subsequent time of detection, as illustrated in **Figure 3**. Through instrument calibration using a set of standards these times can be accurately converted to precise m/z values.

Several important developments in the MALDI-TOF MS instrument design over the past two decades have enhanced the quality (resolution/sensitivity) of data that can be acquired using these instruments. The development and implementation of high-speed computers and electronics have undoubtedly played a major role in these improvements, but the most significant technological
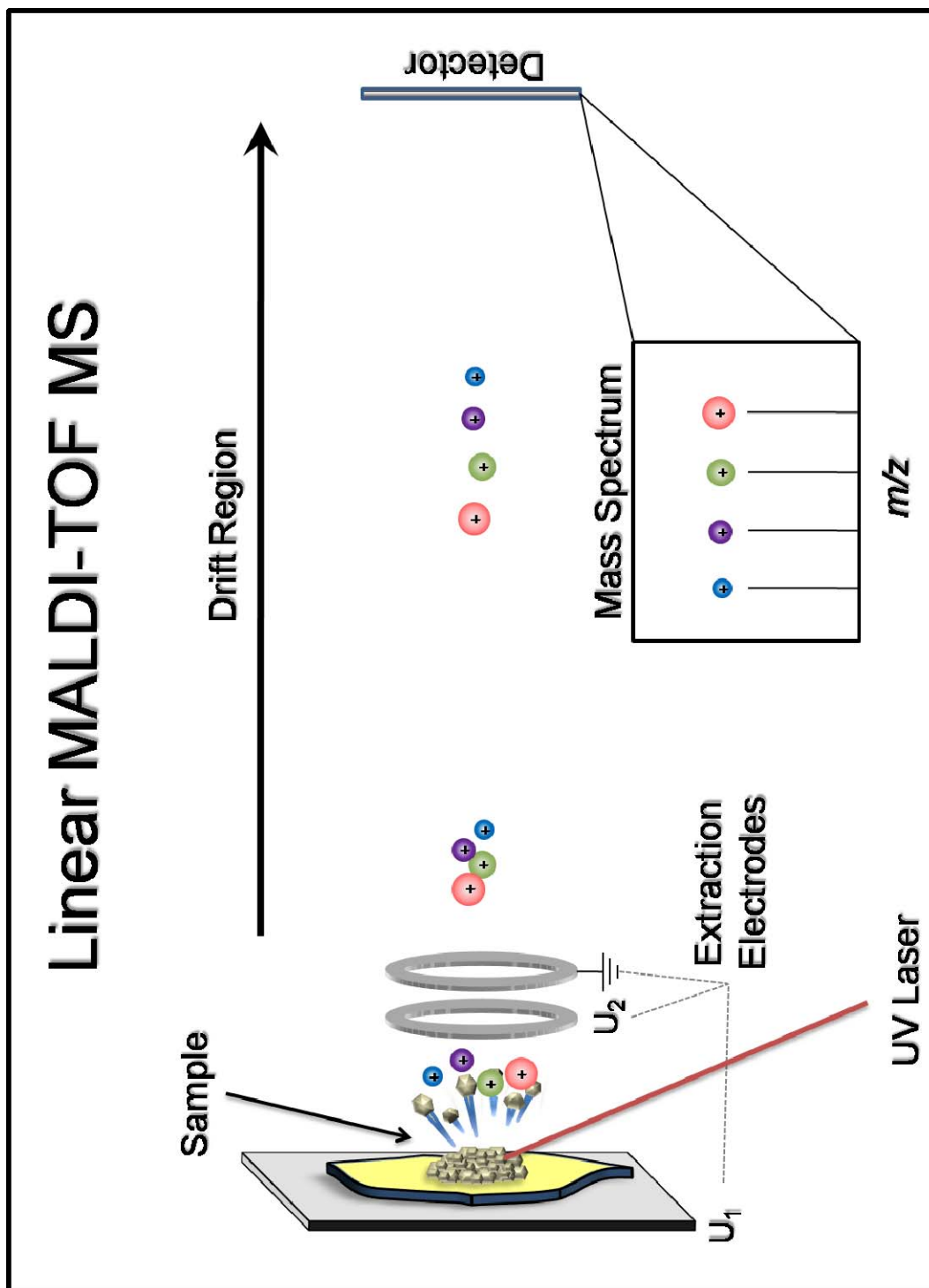
Figure 3. Schematic of a linear MALDI-TOF MS. Following the MALDI process, ions are accelerated into the field-free drift region where they are separated based on their *m/z* values.

advancements are the development of delayed extraction [51-53] and the reflectron (ion mirror)[54].

Delayed extraction functions by introducing a time delay between the ionization and acceleration steps. The desorption process in MALDI results in a distribution spread of the kinetic energies of the resulting ions. When a considering a collection of ions with the same m/z (isobaric), the ions ejected from the sample surface with greater kinetic energy have a higher velocity and therefore travel farther from the sample surface prior to application of the accelerating potential. Those ions ejected with less kinetic energy, and therefore decreased velocity, will be positioned closer to the target surface. Upon application of the accelerating voltage, ions closer to the target surface will experience a greater accelerating potential relative to the ions farther from the surface. With the appropriate delay time, the slower ions will obtain enough additional potential energy to account for the positional discrepancy and as a result all of the ions will be focused at a certain plane in space (such as the detector or entrance to the reflectron). Additionally, delayed extraction allows for the plume (ions + neutrals) to expand prior to extraction therefore reducing the number of collisions that occur between the accelerating ions and the relatively immobile neutral molecules once the voltage is applied.

A reflectron consists of a series of evenly spaced electrodes at the end of the linear flight tube. An electric field gradient of the same polarity as the ions is applied across the reflectron and therefore upon entry into the reflectron, ions are decelerated to zero velocity and then reaccelerated in the opposite direction

towards the reflectron detector.[54] This has a two-fold effect on improving mass resolution: first, the effective length of the flight path is increased, and, second, the flight paths of ions with the same m/z but different kinetic energies are altered such that those with higher kinetic energy penetrate deeper into the reflectron, increasing their flight time relative to the slower ions and therefore focusing the timing at which the ions reach the detector.[54]

For protein imaging (> m/z 3000) the TOF MS is typically operated in linear mode (**Figure 3**), which provides a wide functional mass range of detection and high sensitivity. For lower molecular weight species (< m/z 3000), the TOF MS is operated in reflectron mode (**Figure 4**), which uses an ion mirror, or reflectron, to compensate for the initial velocity/energy distribution and significantly improves mass resolution.

A MALDI-TOF/TOF instrument[55] (Ultraflex III; Bruker Daltonics; Bremen, Germany) can be used to generate tandem MS data on ions directly from a tissue sample. In this instrument, ion fragmentation occurs through a mechanism called post source metastable decay. Ions generated by MALDI can naturally undergo metastable decay through unimolecular decomposition, resulting in single or double backbone cleavages as illustrated in **Figure 5**.[56]

The MALDI-TOF/TOF instrument has the same basic design of a reflectron mass spectrometer, but it uses several additional components to facilitate ion fragmentation and subsequent detection as a tandem MS spectrum.
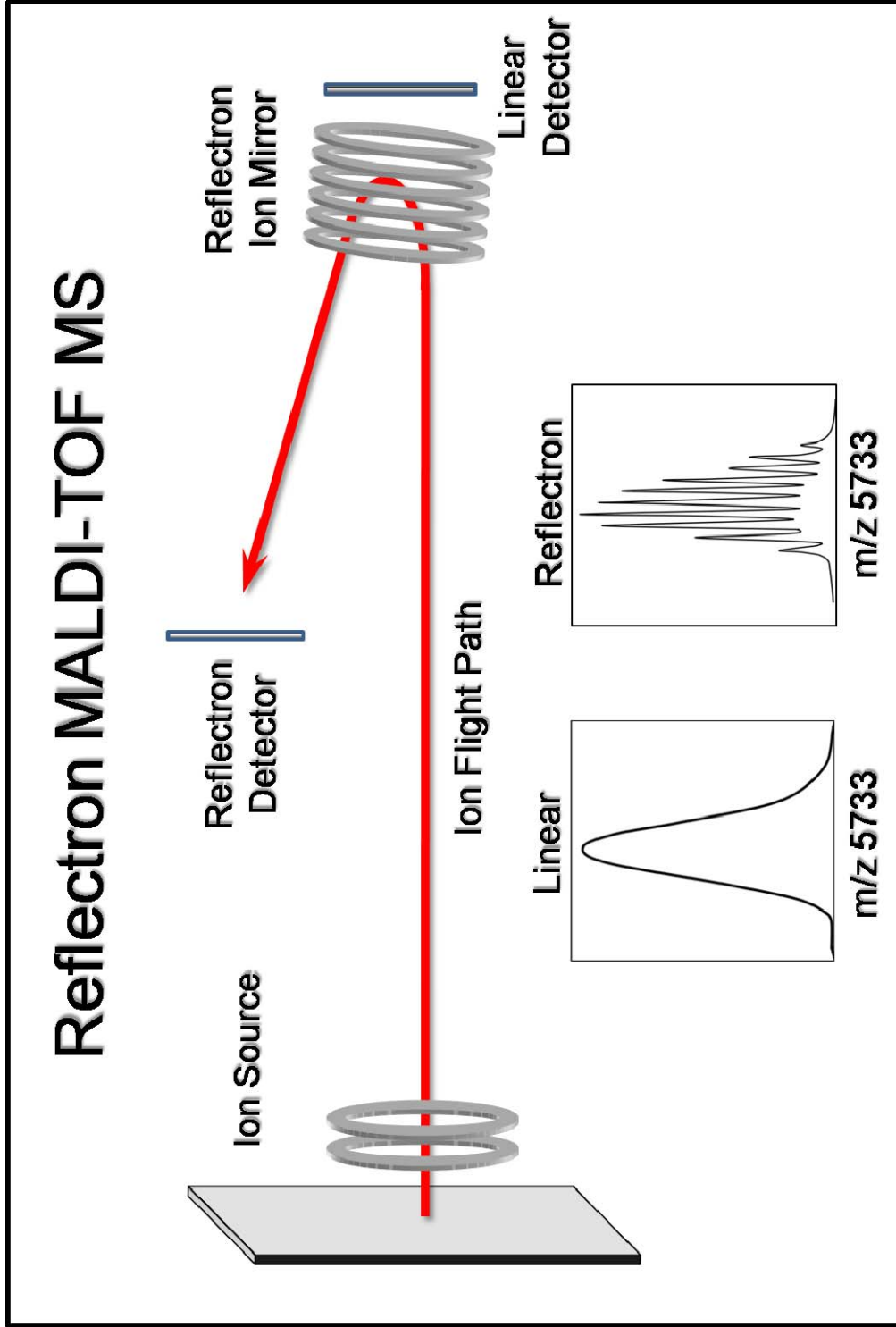
Figure 4. Schematic showing the ion flight path in a reflectron MALDI-TOF MS. The increased mass resolution achieved when the reflectron is used is illustrated for *m/z* 5733.
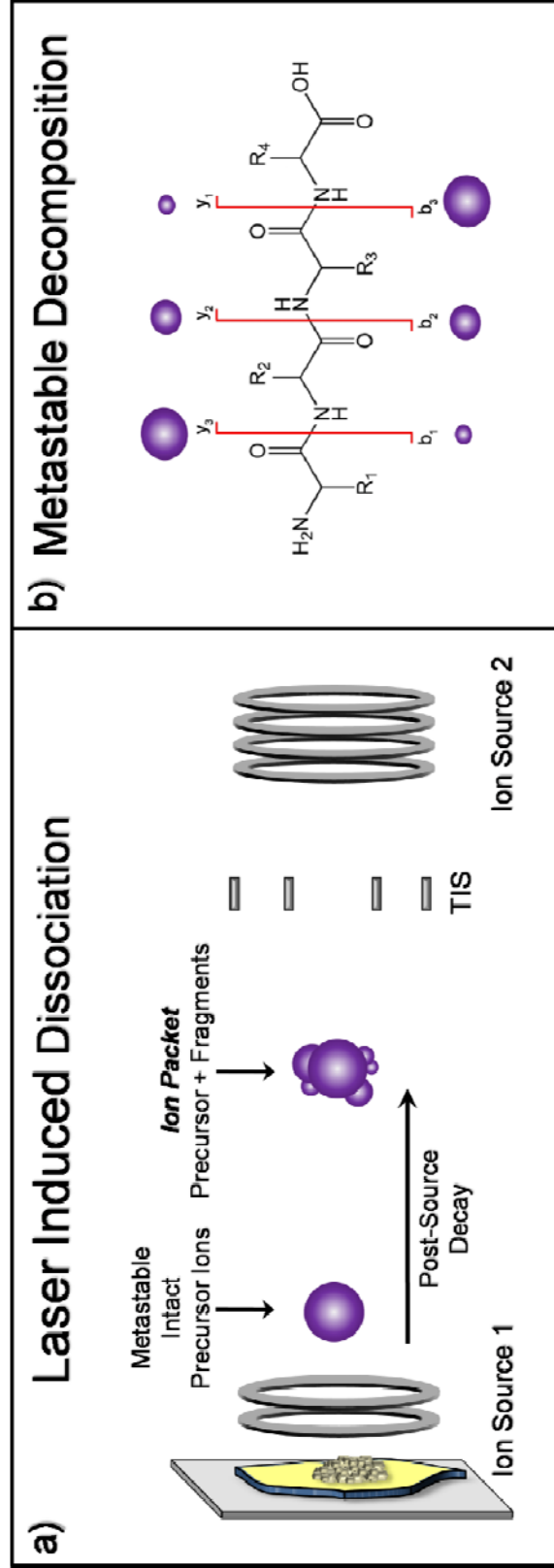
Figure 5. a) Illustration of a metastable precursor ion undergoing post source decay to from fragment ions. **b)** Schematic of the typical peptide background cleavages that take place during metastable decomposition.

Following ion formation and acceleration to 8 kV, all ions have a velocity corresponding to $E=1/2mv^2$. The fragments that form post acceleration through metastable decay maintain the velocity of the corresponding precursor forming an "ion family". A timed ion selector (TIS) is placed in the flight tube and is used to select only the precursor ion of interest, as shown in **Figure 6**. This is achieved by applying a potential across the flight path to deflect all of the passing ions other than the precursor. At the precise timing when the selected precursor ion is present in the TIS the potential difference across the TIS is brought to zero and the ion family continues to traverse the flight tube unimpeded.

The selected ion family leaves the TIS and enters the LIFT device to undergo velocity focusing and further acceleration. The LIFT device is comprised of four electronic grids with 3 stages of operation. In stage 1, the potential energy of the selected ion family is rapidly lifted, hence LIFT, an additional 19 kV. As the ions enter stage 2, which is held at 19 kV as well, they continue to travel at the same velocity and the potential on the third grid is then reduced by 2-3 kV. This accelerates the ions towards Stage 3 where the ions are then accelerated to full speed by dropping the voltage on the fourth grid to ground (analogous to delayed extraction in a regular MALDI ion source) and time-focused onto the detector as illustrated in **Figure 7**.

An additional device called a post lift metastable suppressor (PLMS) is placed between the LIFT cell and the reflectron. The PLMS is similar in design and function to the TIS and serves to deflect the remaining intact precursor ions leaving the LIFT cell. This helps to eliminate undesired fragment ion formation
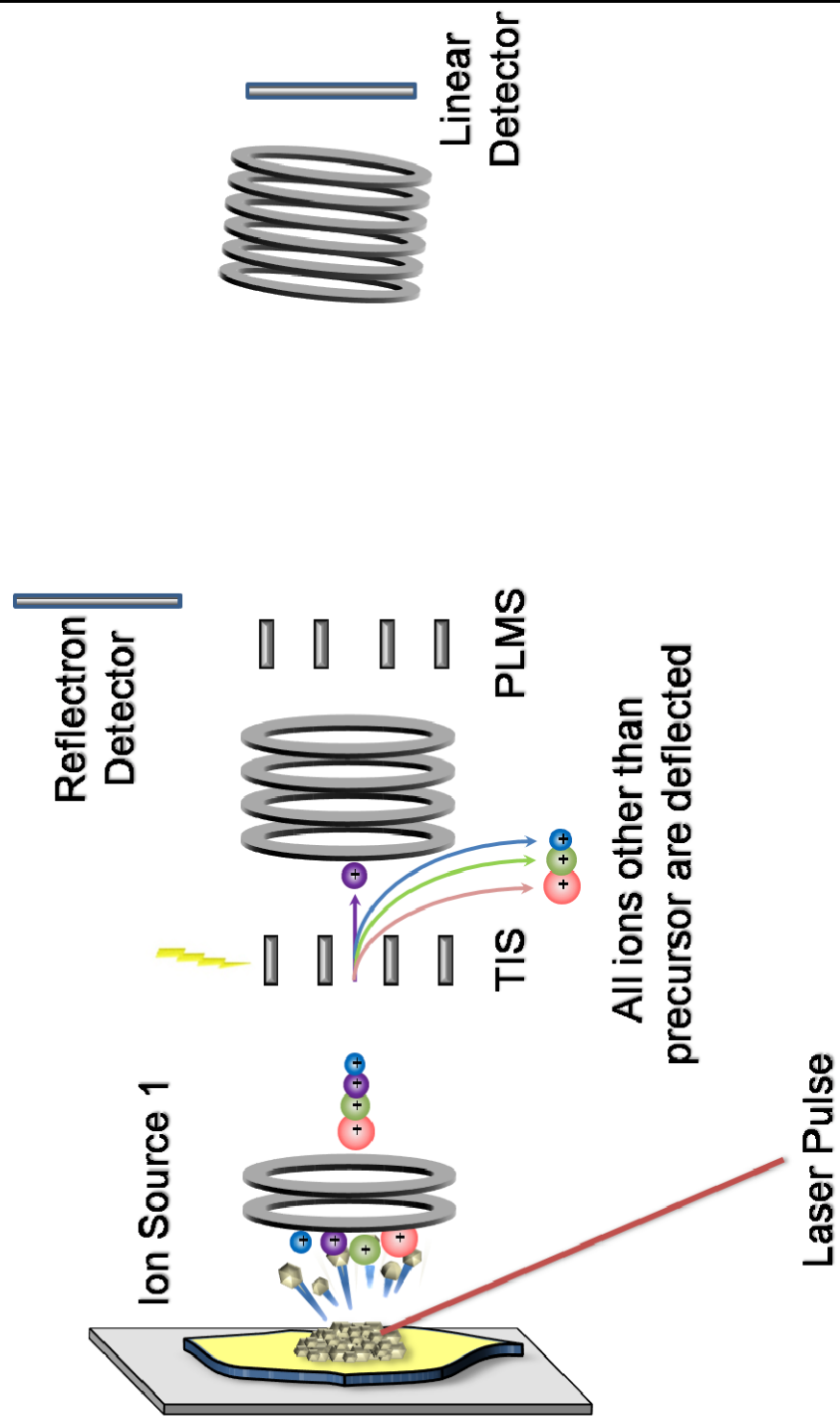
Figure 6. Schematic of the initial steps of a MALDI-TOF/TOF MS experiment where the timed ion selector is used to isolate the precursor ion of interest by deflecting all other ions.
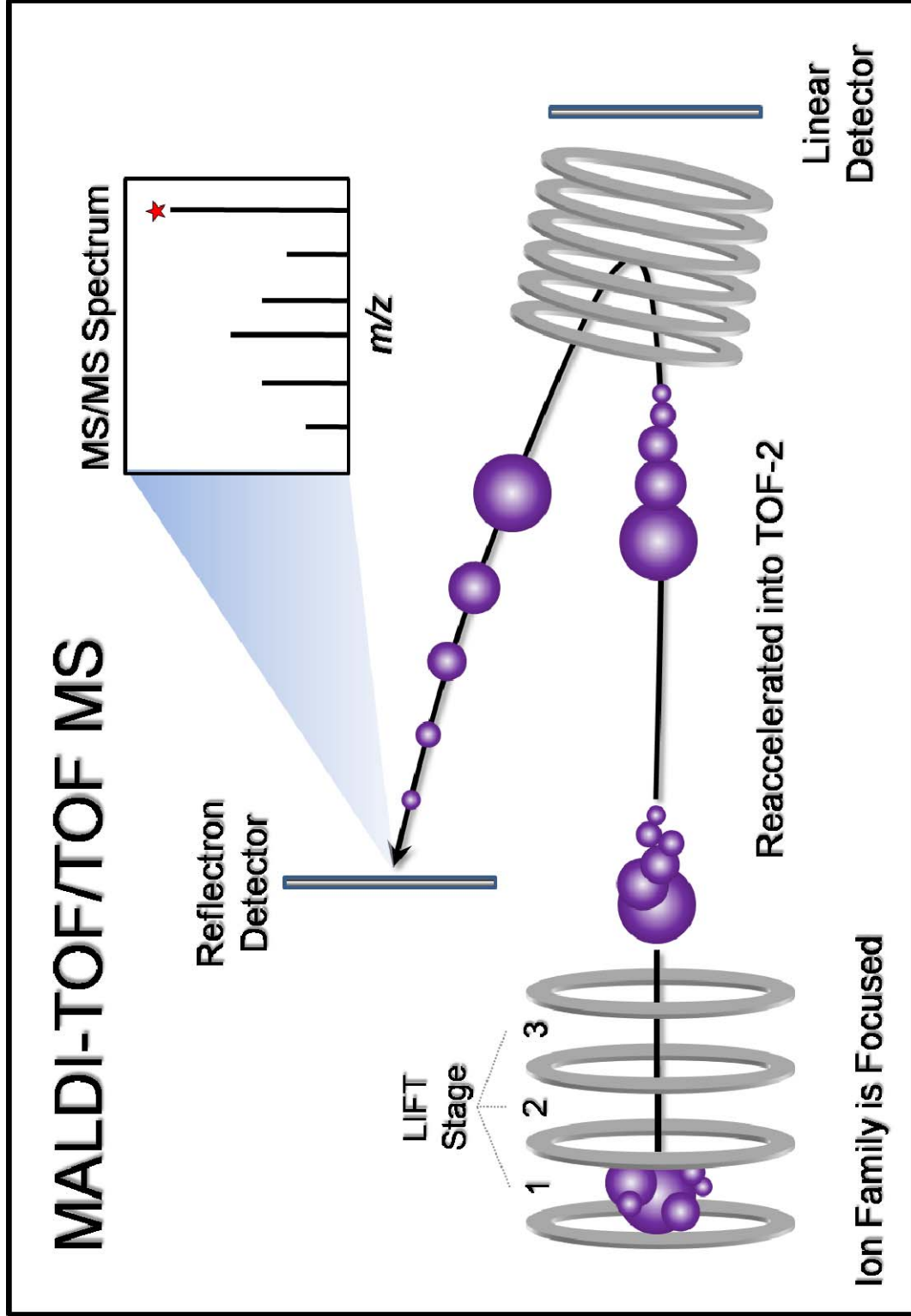
Figure 7. Schematic illustrating the reacceleration of the selected precursor ion family. This enables these ions to be further separated based on their m/z values to generate a MS/MS spectrum.

post-acceleration into the second TOF, which can cause significant chemical background noise in the fragmentation spectrum.

The fragment ion spectra acquired using a MALDI-TOF/TOF resembles those generated using other instruments, such as an ion trap, in that the major ions detected are *a*, *b*, and *y* ions. Interpretation of these spectra is carried out using commercial software which generates a proposed amino acid sequence based on the detected fragments and precursor ion mass. This sequence is then searched against a protein database to determine the representative intact protein.

Data Processing and Analysis

IMS experiments often result in very large datasets consisting of several thousand mass spectra, each of which contains hundreds of peaks. There are several software packages that can efficiently manage IMS data and these programs automate several functions including setup and acquisition, spectral processing, image generation, and statistical analysis.

Following acquisition, several spectral processing steps are typically carried out including baseline subtraction, smoothing, peak picking, and normalization. Baseline subtraction uses an iterative algorithm to remove the baseline slope and offset that is caused mainly by background chemical and electronic noise. Smoothing a mass spectrum averages neighboring data points to increase the signal/noise ratio and enhance the peak shapes. Peak picking algorithms analyze the data points in a spectrum and differentiate peaks

23

corresponding to real analyte ions from those attributed to noise and background. Thresholding parameters, such as a S/N cutoff value, are typically used to control and limit the peak selection process. In reflectron mode, the TOF MS is capable of resolving the isotopes of an individual peptide or lipid species, which naturally result from the presence of one or more ^{13}C atoms (**Figure 4**). Peak picking algorithms, such as the SNAP algorithm developed by Bruker Daltonics, are able to recognize these isotopic clusters and assign a representative mass only to the monoisotopic peak, which is the peak from the molecule containing only ^{12}C atoms and no ^{13}C atoms. This can be an extremely important step in processing peptide and lipid spectra, because it results in a significant reduction in data complexity. For statistical analysis, monoisotopic peak picking generates a single peak to represent each biomolecule, as opposed to several isotopes, and can be used as the comparative value.

Another important aspect of data processing is spectra normalization. This is typically carried out by evaluating the total ion current (TIC) of all spectra in a dataset. The TIC is defined as the sum of the intensities of all data points in a spectrum. Notwithstanding certain assumptions, it is expected that this value should be similar between all spectra in a dataset. Deviations in the TIC likely result from slight day-to-day fluctuations in instrument performance, variations in sample preparation, and differing chemical microenvironments in tissues. When normalizing on a spectrum to spectrum basis, all data points are adjusted by the same increment so that the resulting TIC of each independent spectrum is equal to a normalized value.

Other normalization functions are also used, including the division of each data point in a spectrum by the TIC of the entire spectrum to bring all data point values within the range of [0,1]. This type of normalization is independent of the other spectra in a dataset, because each data point is essentially normalized based on its relative contribution to the TIC.

Biological Applications of IMS

Over the past decade IMS and direct tissue profiling experiments have been conducted on a wide variety of biological samples and have provided investigators with valuable insight into the underlying biology. IMS has the ability to provide a comprehensive view of a subset of a tissues proteome in an extraordinarily short amount of time. As a result, examining the differences between a set of normal and diseased tissues is a particularly useful application of this technology. For example, IMS has been used to examine the proteomic changes involved in the development and progression of numerous cancers including gliomas[13, 57, 58], breast cancer[59, 60], prostate cancer[61, 62], colon cancer[63], and lung cancer[64, 65]. These experiments revealed numerous differential protein expression patterns between cancerous and normal tissues.

Recently, Caldwell et al.[66] examined soft tissue sarcomas using IMS and determined that tissue adjacent to the tumor that appeared histologically normal, actually exhibited changes at the molecular level consistent with the cancerous regions. These tumor associated proteins showed a gradient effect across the histologically defined tumor margin into the normal tissue at distances up to 1.5

cm. A larger study that performed a similar analysis on tumors from patients with clear cell renal cell carcinoma showed comparable results (Stacey R. Oppenheimer, Unpublished Data). Statistical analysis of the data revealed several interesting patterns, including proteins that dropped off sharply at the margin, those that extended beyond the tumor margin into the normal, and those that appeared only in the normal. These results demonstrate that changes are likely taking place at the molecular level that cannot be detected using histopathological analysis alone. The implications of this are important not only from a cancer biology perspective, but also in the context of surgical removal of tumors. For example, during the surgical removal of tumors it is imperative that all cancerous cells are removed in their entirety to improve a patient's chance of survival and prevent local recurrence. This data suggests that cancer cells may be invasive beyond the histological tumor margin and could likely play a role in tumor recurrence at these sites.

Other research has focused on the use of mass spectrometry data acquired through imaging and profiling studies to differentiate and classify different tissue histologies. For histological classification, pathologists typically use light microscopy to examine cellular morphology and various antibody stains to assess the presence or absence of histology specific proteins. The advantage of incorporating IMS into this workflow is that the spatial distribution of hundreds of protein species can be visualized simultaneously, adding an extraordinary amount of additional information.

Histology directed imaging and profiling refers to the application of IMS to the molecular analysis of targeted regions of clinical samples. In this workflow, a pathologist reviews a tissue section serial to the section used for MS analysis to determine specific areas of interest.[67] In-house software was developed to guide the MS analysis to only acquire mass spectra from these pathologically relevant regions, which can significantly reduce the overall time of the MS analysis by analyzing only specific regions of interest.[67] Furthermore, each mass spectrum can be directly associated with a specific histological region.

Rahman et al. applied this workflow to lung cancer tissues and were able to classify normal epithelium, pre-invasive, and invasive lung lesions with 90% accuracy based on proteomic differences.[68] Additionally, Schwartz et al. showed that this technology could effectively classify the grade of different human gliomas and predict survival of the patients.[69]

Summary and Research Objectives

Biological systems are extremely complex and new tools are necessary to effectively enhance our knowledge and understanding of their underlying mechanisms. Mass spectrometry has proven to be an essential technology for the analysis of biological samples due to its unparalleled throughput, sensitivity, and specificity. IMS combines the power of MS with the inherent spatial information contained in a biological sample revealing biomolecular distributions on a scale unattainable by other techniques.

The work presented here has two main focuses. First, chapter 2 details the development of a method to carry out *in situ* chemical reactions to enhance the proteomic information that can be detected in a tissue sample. This method, referred to as *in situ* digestion, uses an enzyme to digest the proteins present in discrete regions of a tissue prior to matrix application and MS analysis. The enzymatically digested proteins result in a large collection of proteolytic peptides in the range of m/z 700-3000. This collection of peptides can be sequenced and identified directly from the tissue using tandem MS and then linked back to their respective intact protein. The spatial information for these ions can be used to validate their identification because all peptides from the same protein should exhibit the same spatial distribution. *In situ* digestion also enables the indirect detection of high molecular weight proteins (> m/z 30,000) that are not detected in a standard protein imaging experiment by mapping the peptides generated from these species. Finally, this method enables the analysis of formalin-fixed paraffin-embedded (FFPE) tissues. FFPE tissues are a valuable and extensive source of clinical samples which are not compatible for use in standard protein imaging experiments.

Second, chapters 3 and 4 detail the application of this method to FFPE lung cancer tissue microarrays (TMA). We show that *in situ* digestion coupled with imaging mass spectrometry can be used to reproducibly map numerous protein species in a large population of lung cancer patients. Several of these proteins correspond to previously reported markers of lung cancer, while others have not yet been associated with this disease. Furthermore, we discovered

patterns of protein expression that can be used to accurately distinguish and classify the histological subtypes present in the sample cohort.

In summary, this work shows that when IMS is performed on a large set of tissues in parallel with a pathological evaluation, the combined histological and molecular information provides a depth of information that has the potential to revolutionize disease diagnosis, prediction of prognosis, and course of therapy.

CHAPTER II

IDENTIFICATION OF PROTEINS DIRECTLY FROM TISSUE: *IN SITU* TRYPTIC DIGESTIONS COUPLED WITH IMAGING MASS SPECTROMETRY

Abstract

A novel method for on-tissue identification of proteins in spatially discrete regions is described using tryptic digestion followed by MALDI Imaging Mass Spectrometry (IMS) with MS/MS analysis. IMS is first used to reveal the protein and peptide spatial distribution in a tissue section and then a serial section is robotically spotted with small volumes of trypsin solution to carry out *in situ* protease digestion. After hydrolysis, matrix solution is applied to the digested spots with subsequent analysis by IMS to reveal the spatial distribution of the various tryptic fragments. Sequence determination of the tryptic fragments is performed using MALDI MS/MS analysis on-tissue directly from the individual digest spots. This protocol enables protein identification directly from tissue while preserving the spatial integrity of the tissue sample. The procedure is demonstrated with the identification of several proteins in coronal sections of a rat brain.

Introduction

Desorption of proteins and subsequent analysis by MALDI MS provides a measurement of molecular weight, but not the identification of the protein.

Protein databases do not generally include the sometimes extensive post-translational modifications of most proteins: indeed such modifications are still unknown in most cases. Further, nominal molecular weight redundancies to within the mass measurement accuracy of the experiment can be problematic. Current protein identification techniques are based on extraction, proteolysis, and LC/MS/MS techniques. Although these procedures are effective, they can be tedious, require relatively large amounts of sample, and result in loss of spatial information. Nonetheless, this procedure remains an effective means of identification of an unknown peak in a spectrum. Several recent papers have attempted to address some of these problems through automation. For example, a molecular scanner approach involving tissue blotting onto a trypsin membrane, followed by capture of the resulting peptides onto a second membrane can be used to digest proteins while maintaining some degree of spatial resolution [70-72]. In addition, several studies [73-82] have reported MS/MS data directly obtained from tissue samples.

In the current work, we show that automated deposition of a trypsin solution can be effectively used to carry out digestion of proteins and peptides directly on thin tissue sections in well-defined micro-spotted arrays.[83] These arrays can then be automatically spotted with matrix solution for subsequent MALDI MS and MALDI MS/MS analyses to obtain sequence information for the tryptic peptides and thereby enable protein identification. The array printer used for reagent deposition ensures accurate and precise droplet placement to enable well-controlled and reproducible digestion conditions and is an essential part of

the process. Confining the digestion enzyme to within discrete spots on the order of ~200 μm in diameter reduces the complexity of the analysis because the proteins digested are representative of only a limited number of cells and extracellular space. Spotted arrays have been found to be much more reproducible and compatible with automation than spray coating and other matrix deposition techniques [84]. Individual droplet placement allows for a micro-extraction process to occur on the tissue surface at each position over several deposition iterations. This enhances the efficiency of the protein digestion and subsequent matrix crystallization process, thereby improving sensitivity and signal quality. The bilateral symmetry of the coronal brain sections used in this report provide a measure of the reproducibility of the spot-to-spot digestion efficiency, since the relative intensity of the tryptic peptides generated from the proteins present in the substructures of each hemisphere should be equivalent. The novel aspect of this technique lies in its capability to not only visualize, but to also identify proteins from discrete areas of a tissue section encompassing areas of about 200 μm diameter. This technique is optimal for the verification of protein identities where specific proteins are suspected to be present. In these cases, the digest spot can be searched for known tryptic fragments of a protein and the peptide sequenced by MS/MS if found. The ability to robotically carry out controlled and reproducible trypsin digestions within thousands of discrete spots throughout a tissue section enables positive identification of numerous protein species with high spatial localization and a relatively small amount of sample preparation and time requirement.

The advantage of this method over traditional protein visualization techniques, such as immunohistochemical staining, is that in a single analysis it is possible to examine the distribution of numerous proteins and peptides while simultaneously obtaining *in situ* identification. The digest data can be examined in parallel with the information generated by the peptide and protein imaging experiments to provide a comprehensive analysis of a tissue section. This approach is ideal for discovery since it does not require specific reagents for molecular detection (such as an antibody).

Results

The spectra generated at each spot on the digested tissue typically contain several hundred distinct peaks with a S/N > 5 (**Figure 8**). The signal intensity of the tryptic peptides vary widely due to several factors, i.e. different expression levels of proteins, variations in digestion efficiency, and differences in desorption and ionization efficiencies of different peptides. Nevertheless, clearly resolved molecular species are easily identified in these digest spectra as shown in the inset in the upper left of **Figure 8**. Myelin basic protein (MBP), which is expressed regionally in the brain at relatively high levels and is essential for the formation of central nervous system (CNS) myelin and neuronal transmission [85], was among the proteins identified using the workflow shown in **Figure 9**. There are several molecular forms of MBP expressed in the CNS including a number of splice variants and numerous post translational modifications [85]. The major isoform in adult rats has been reported to have a molecular weight of 14.2 kDa [85].
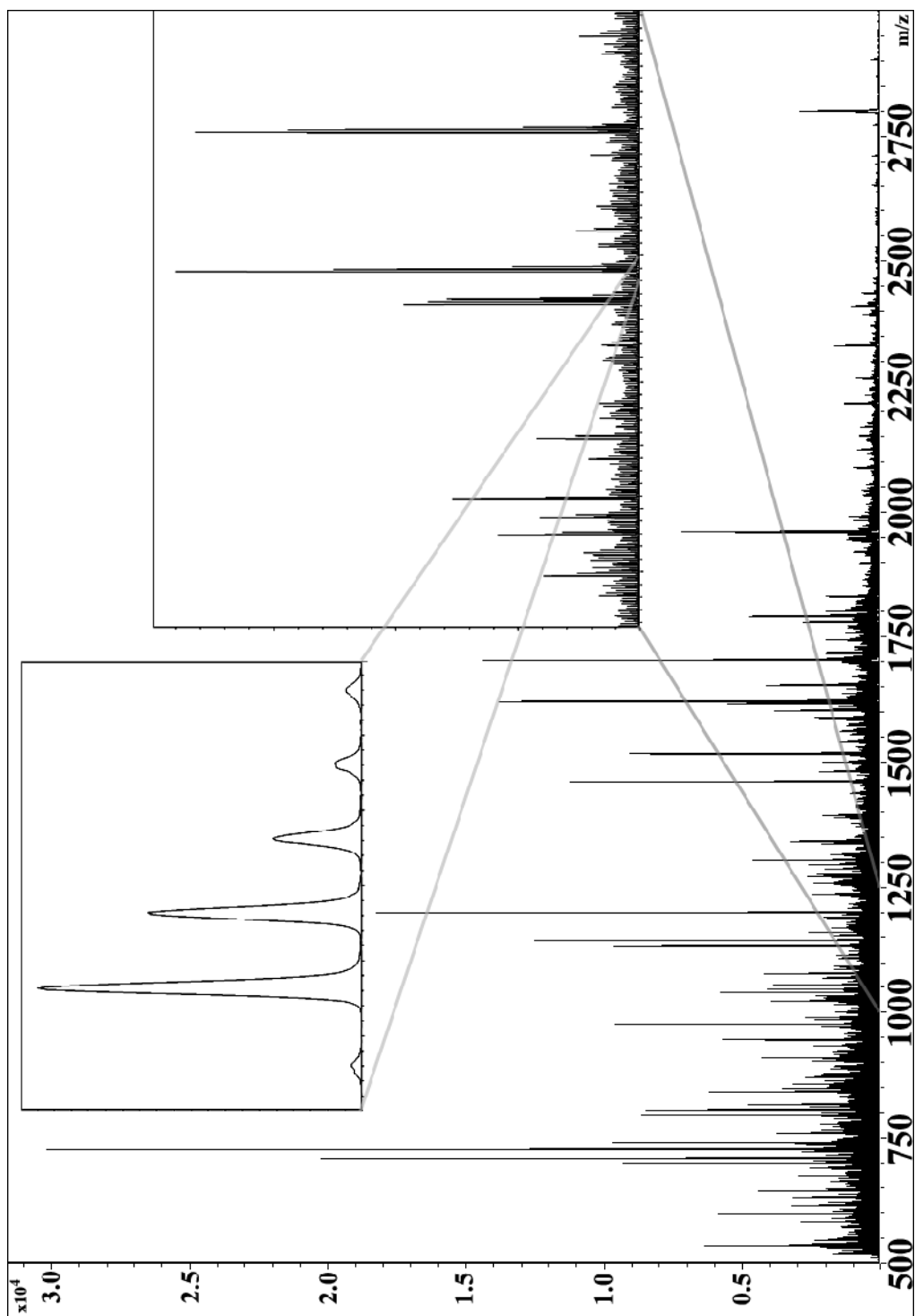
Figure 8. Representative spectrum from an individual digest spot.

34
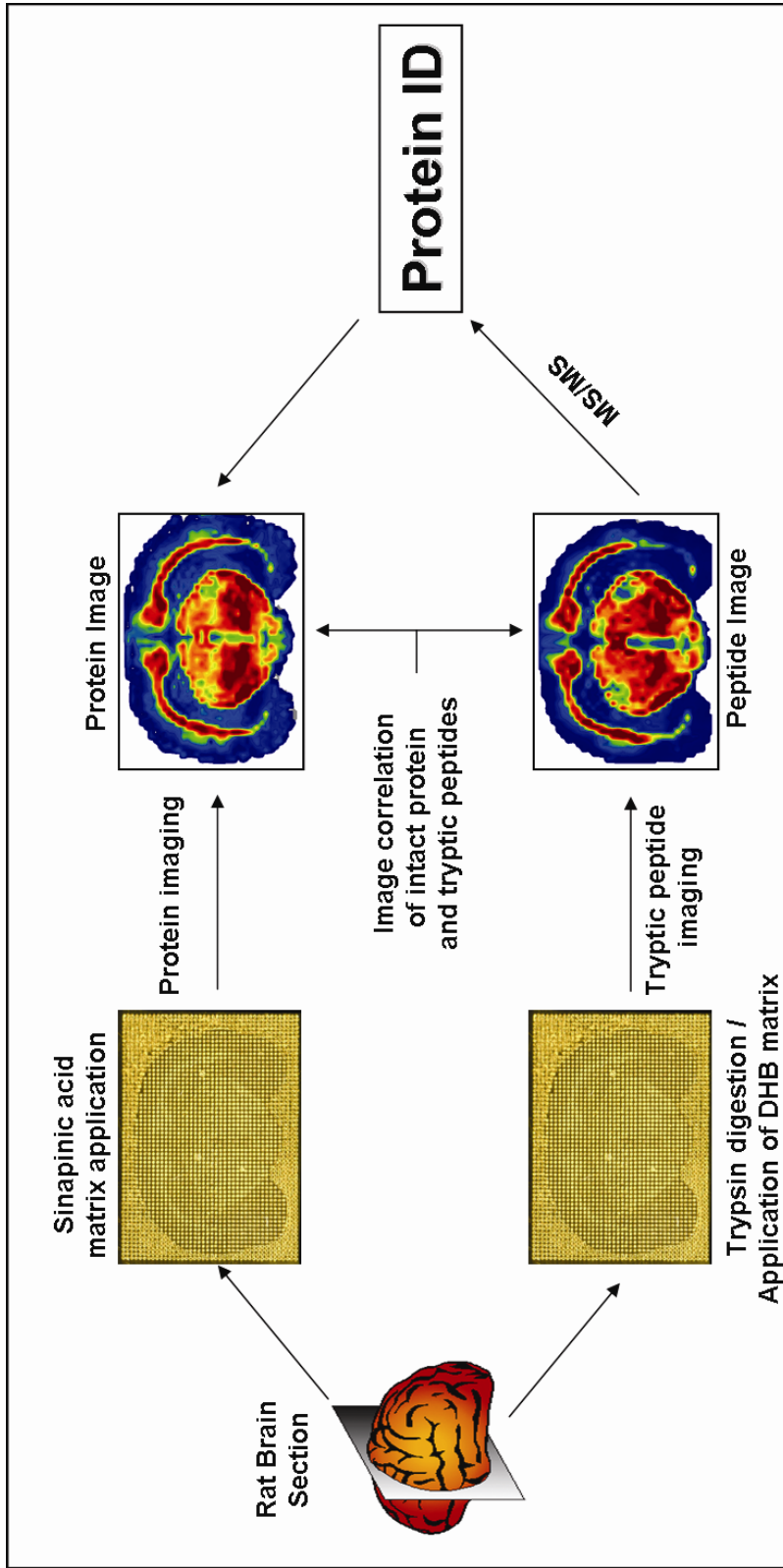
Figure 9. Workflow for on-tissue protein identification

The distribution of this intact protein was determined using IMS on a coronal section of rat brain tissue as shown in **Figure 10**. After digestion, a total of 8 tryptic peptides from MBP were detected (**Figure 11**) and sequenced directly from the tissue (**Figure 12**). Other tryptic peptides were also generated and detected from MBP, although it was not necessary to sequence them since there was already sufficient data for a high confidence identification of this protein. Proteins identified in this way were further validated by comparison of the spatial distribution of a given parent protein with sequenced peptides generated from that protein. Internal calibration of the acquired spectra using the masses of trypsin autolysis fragments as well as several other known tryptic peptides sequenced with high confidence through MS/MS from proteins present throughout the rat brain were used to achieve mass accuracies of about 10 ppm. As expected, the MBP tryptic peptides show an ion density distribution consistent with that of the intact protein. It is noted that other isoforms of MBP were present in the protein analysis at lower intensity.

As further examples, other proteins identified in the same process from a different region of a rat brain include neurogranin, a 7496 Da brain-specific protein kinase C (PKC) substrate [86] and brain-specific polypeptide (PEP-19), a 6676 Da neuron specific protein [87] **(Figure 13)**. Each of the four tryptic peptides displayed for each protein were sequenced by MS/MS and linked to their respective parent protein through database matching and spatial distribution correlation. Neurogranin was distributed with highest intensity throughout the cerebral cortex and with a lesser intensity in the striatum, which is in accordance
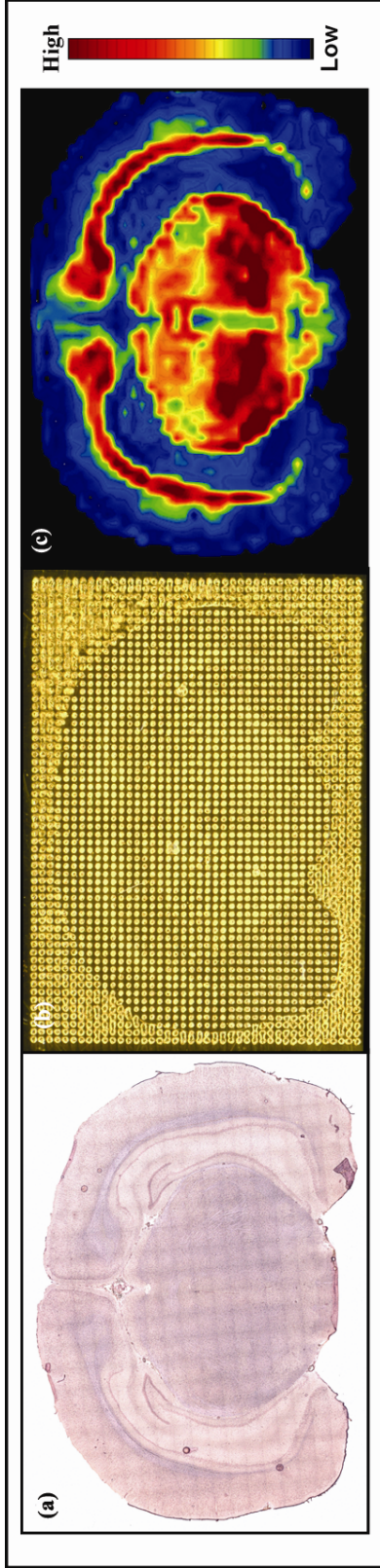
Figure 10. a) H&E stain of rat brain tissue section serial to the sections used for digestion and imaging b) Tissue section spotted with a sinapinic acid matrix solution c) Image of the 14.2 kDa isoform of myelin basic protein
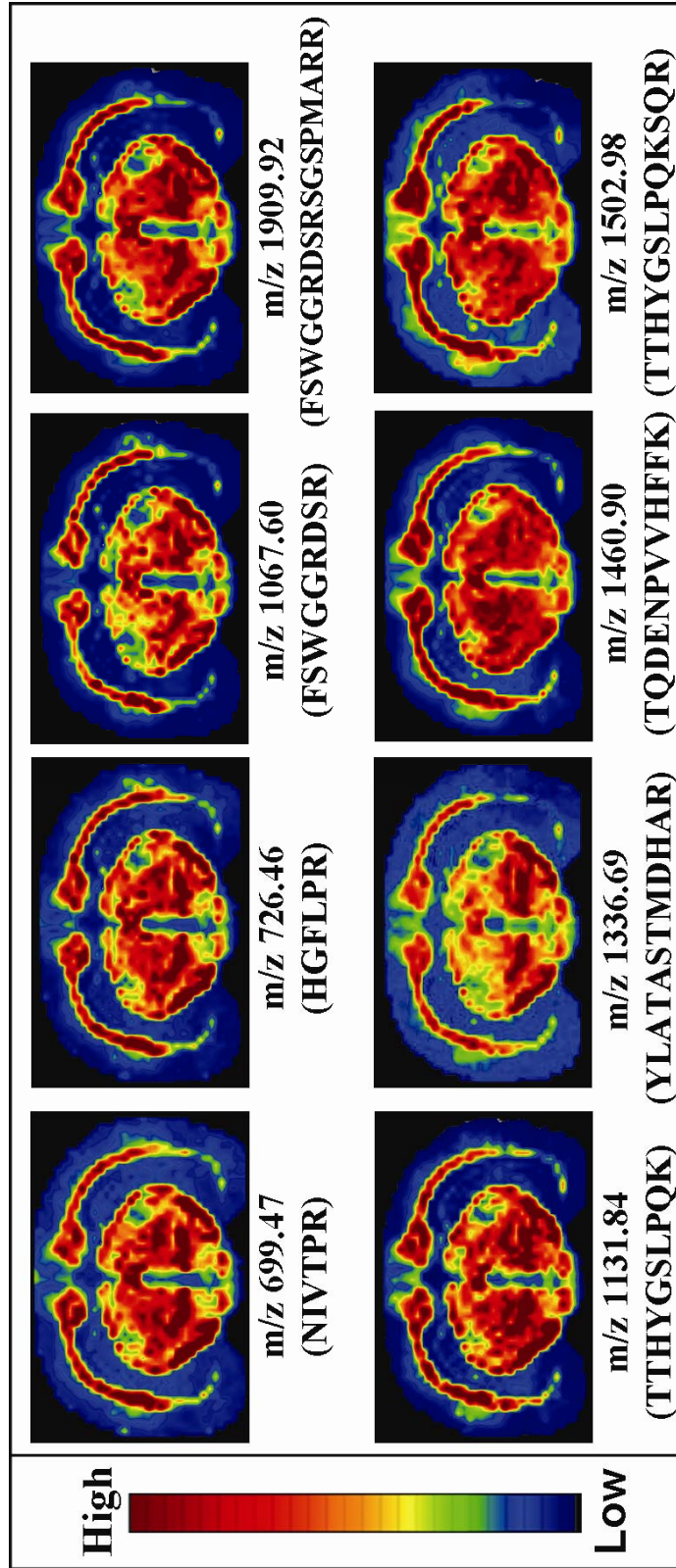
Figure 11. Images of tryptic peptides generated from the digestion of the 14.2 kDa isoform of Myelin Basic Protein
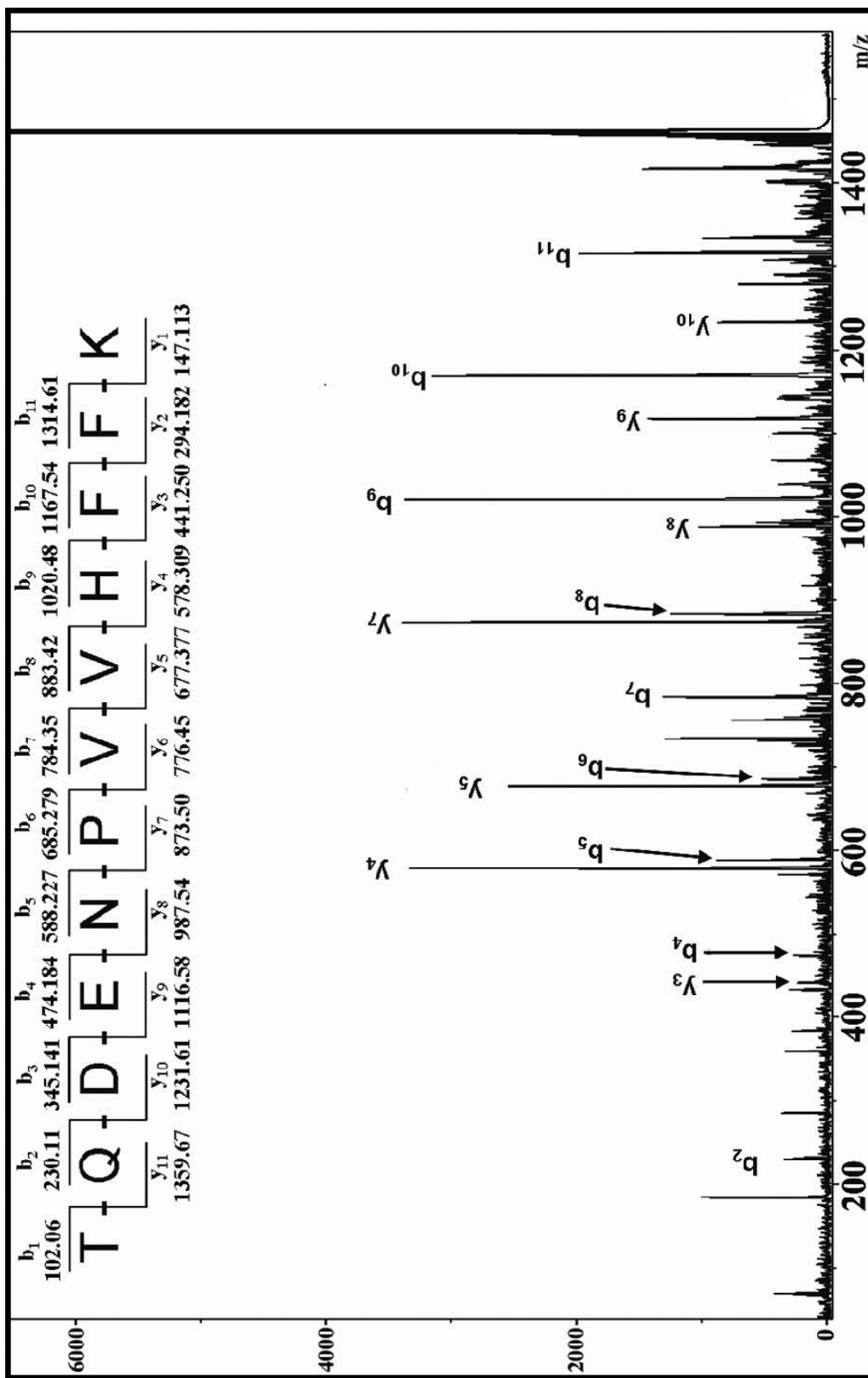
Figure 12. MS/MS spectra of ion at m/z 1460.90 acquired directly from tissue section following digestion and matrix coverage.
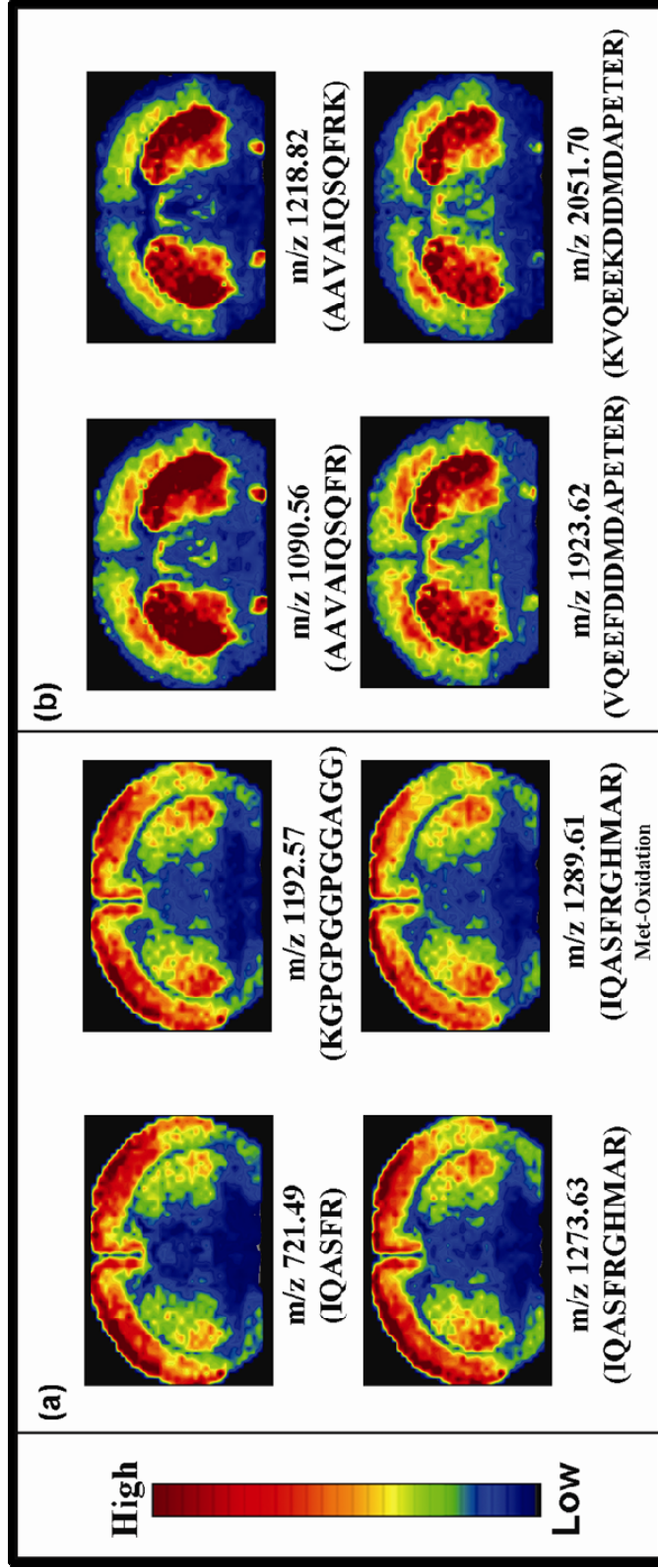
Figure 13. (a) Tryptic peptides generated from the digestion of the 7.5 kDa protein Neurogranin. (b) Tryptic peptides generated from the digestion of the 6.7 kDa protein PEP-19.

with the findings of Represa et al. [86] who examined the distribution of this protein in the adult rat brain through immunohistochemistry. PEP-19 is shown to be distributed with highest intensity throughout the striatum, and with lower intensities in the globus pallidus. These findings also correlate well with previous studies that have determined, through immunohistochemistry, that PEP-19 is highly expressed in the basal ganglia [87]. These data further demonstrate the ability of this method to confidently identify proteins using both digest and MS/MS data in parallel with IMS. The data obtained for the identification of these three proteins is summarized in **Table 1**.

An initial step in these experiments involves determination of peptides present throughout the tissue prior to digestion. This not only provides a higher level of confidence in assessing those peaks in the digest spectra as being generated from proteolysis, but also provides information about the presence of endogenous peptides in different regions of a tissue. For example, the structure of the globus pallidus, which is a major element of the basal ganglia system, can be distinguished in the native peptide image from several ion density maps of the *in vivo* processing products of the proenkephalin A precursor into its propeptide and active peptide forms (**Figure 14**). The propeptide corresponding to the 143-185 region of this protein is detected in the non-digest analysis at m/z 4594. Digestion of this peptide generates several tryptic fragments including m/z 2619.24 and 2818.37 which are detected in the digest images and lie in a mass range more amenable to MS/MS analysis. Several of these peptides do not contain lysine or arginine residues and therefore do not undergo digestion from

Table 1. Data for several proteins detected in rat brain sections, giving the MW for detected protease fragments, and the sequence determined from the database search after MS/MS analysis, with M_{exp} representing the experimental mass and M_{calc} the calculated mass. Each of the calculated masses for the proteins includes an N terminus acetylation in accordance with the sequence found in Swiss-Prot.

Measured Protein In Tissue	Tryptic Peptides from *In Situ* Digest				
	Observed m/z	M_{exp}	M_{calc}	ppm	Sequence
Myelin Basic Protein					
Observed m/z: 14124	699.412	698.405	698.408	3.2	NIVTPR
M_{exp}: 14123	726.403	725.396	725.397	1.8	HGFLPR
M_{calc}: 14122	1067.516	1066.508	1066.494	13.0	FSWGGRDSR
	1131.587	1130.579	1130.572	6.6	TTHYGSLPQK
	1336.640	1335.633	1335.624	6.4	YLATASTMDHAR
	1460.701	1459.694	1459.710	10.5	TQDENPVVHFFK
	1502.760	1501.753	1501.764	7.1	TTHYGSLPQKSQR
	1909.929	1908.922	1908.913	4.9	FSWGGRDSRSGSPMARR
PEP-19					
Observed m/z: 6720	1090.609	1089.602	1089.593	8.2	AAVAIQSQFR
M_{exp}: 6719	1218.706	1217.699	1217.688	8.6	AAVAIQSQFRK
M_{calc}: 6719	1923.840	1922.832	1922.820	6.2	VQEEFDIDMDAPETER
	2051.926	2050.919	2050.915	1.9	KVQEEFDIDMDAPETER
Neurogranin					
Observed m/z: 7538	721.394	720.387	720.392	6.8	IQASFR
M_{exp}: 7537	1192.624	1191.616	1191.611	4.6	KGPGPGGPGGAGGAR
M_{calc}: 7538	1273.672	1272.664	1272.651	10.6	IQASFRGHMAR
	1289.681	1288.673	1288.646	21.3	IQASFRGHMAR (Met-oxidaition)
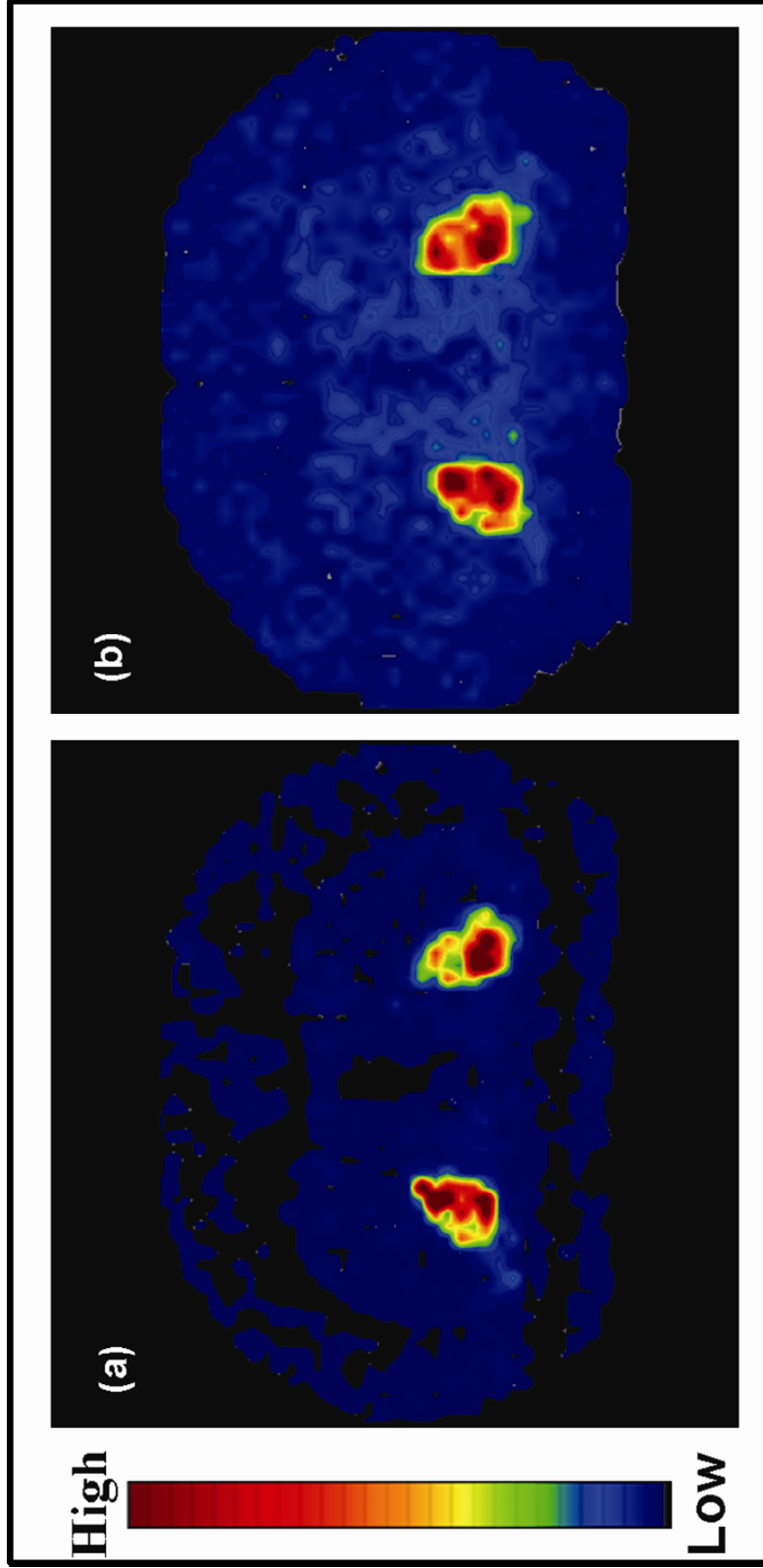
Figure 14. Ion distribution of peptides in the globus pallidus generated from the endogenous processing of proenkephalin A precursor (a), m/z 4595, and trypsin digestion of this peptide (b), m/z 2619.24.

the trypsin enzyme. Consequently, although these peaks are detected in the digest spectra, they can be categorized as non-tryptic peptides.

Signals corresponding to only a few pixels in the ion maps can be used to characterize species localized to isolated regions of a tissue section. For example, *Arg*-vasopressin and copeptin, two peptide hormones generated from the processing of the vasopressin-neurophysin 2-copeptin precursor, were detected solely in the supraoptic nucleus, which is consistent with previous studies [88]. Mass spectra acquired at this position in the non-digest analysis indicate that this precursor is fully processed in this region due to the presence of the active form of *Arg*-vasopressin as shown in **Figure 15.** The calculated mass of Arg-vasopressin is 1086.44, yet experimentally it is detected in this region with a mass of 1083.44. As previously described [89], this mass shift suggests that the C-terminus glycine is amidated and the disulfide bond is intact, the latter of which was further confirmed in the present study by MS/MS analysis. The calculated mass of copeptin is 4281.80 Da, yet it has been shown that this peptide is N-glycosylated [90-92] at an asparagine residue. Two tryptic peptides found at m/z 1798.96 and 2061.09 corresponded to the 21-37 and 21-39 regions, respectively, of the non-glycosylated segment of the copeptin sequence. These peptides were sequenced by MS/MS on a digested tissue section to confirm the sequence. Ivell et al. used tritium labeled sugars to determine that the glycosylation side chain on copeptin contained mannose, glucosamine, and fucose [93]. A peak found at m/z 5930 in the non-digested image is localized to the supraoptic nucleus, identical to that of *Arg*-vasopressin. This corresponds to a mass shift of 1648.5
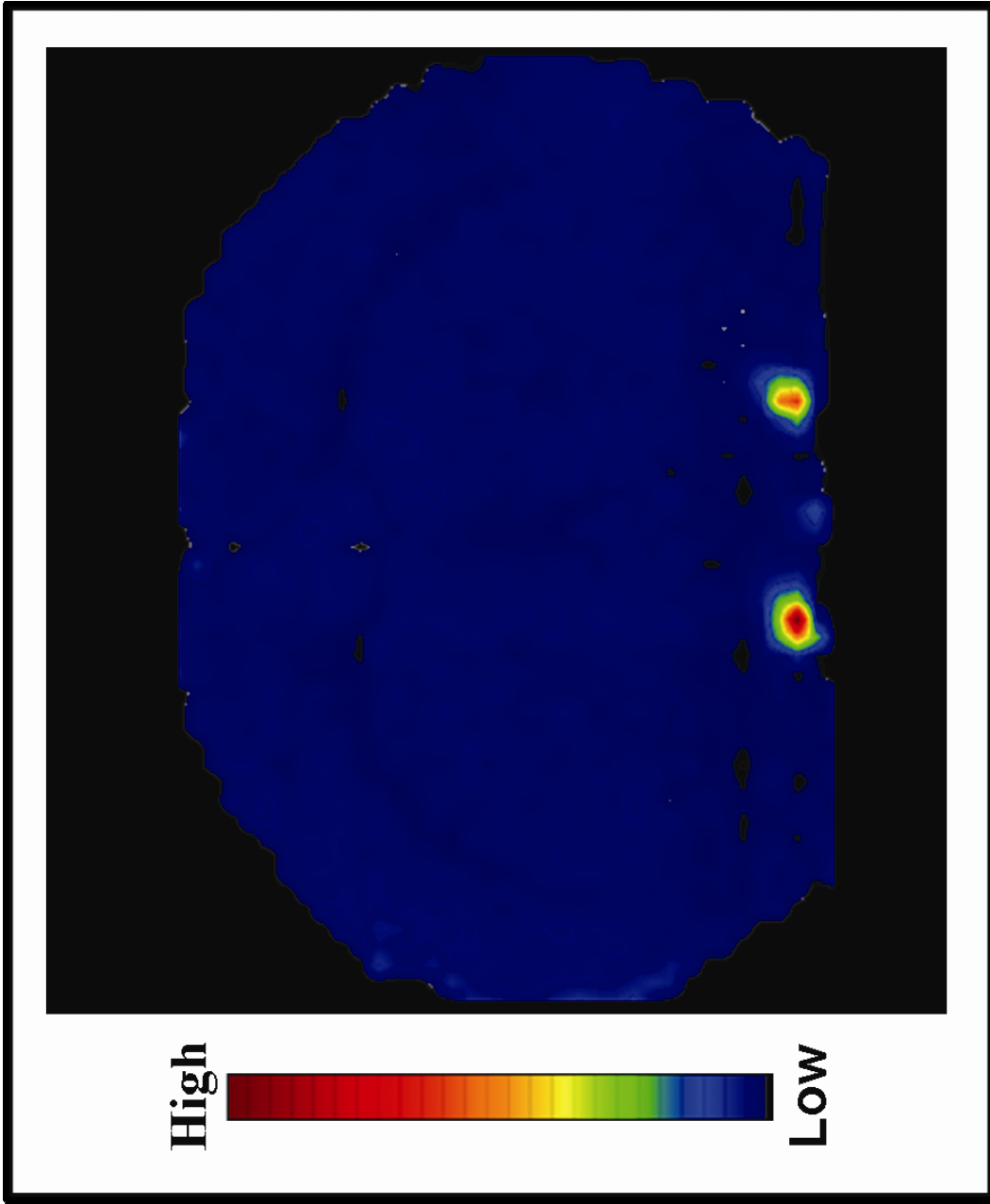
Figure 15. Image generated from vasopressin in the non-digest experiment

Da from the calculated mass of copeptin, which is consistent with a carbohydrate chain composed of $Man_3GlcNAc_5Fuc_1$. This mass shift is also apparent in the digest analysis on the copeptin tryptic peptides consisting of residues 1-20 and 1-15, both of which contain the glycosylated residue. Application of trypsin followed by treatment with peptide N-Glycosidase F (PNGase F) to the posterior region of a rat pituitary section where copeptin is stored [94], resulted in removal of this polysaccharide chain and the tryptic peptides from residues 1-20 and 1-15 were detected at their expected masses. Intact copeptin was also detected at its expected mass in a section treated only with PNGase F.

Tryptic peptides from the digestion of higher molecular weight proteins including actin, a 41 kDa globular structural protein, tubulin, a 55 kDa protein which dimerizes to form microtubules, and synapsin-1, a 74 kDa neuronal protein were identified using this method **(Figure 16)**. Generally, low abundance proteins within the m/z range of these relatively high molecular weight species are difficult to analyze using MALDI-TOF MS. Proteolysis of these high molecular weight proteins directly on tissue and image analysis of unique peptides enables visualization of their distribution throughout a tissue section.

Discussion

MALDI IMS is a unique tool for analyzing the spatial distribution of peptides and proteins throughout tissue sections, providing an enormous amount of data with minimal sample preparation time. Molecular analysis directly on-tissue for identification of specific proteins offers an alternative to the time

46
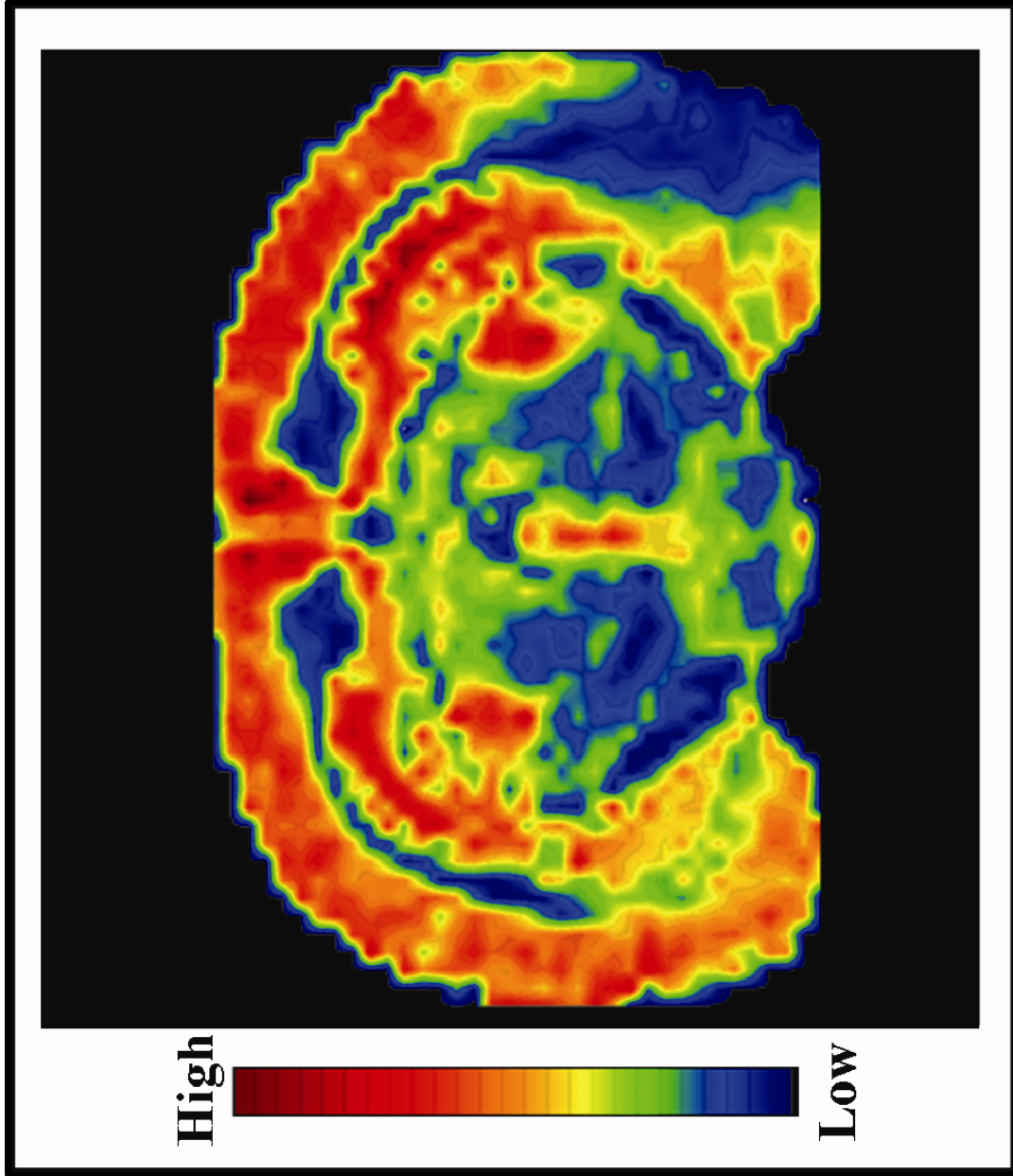
Figure 16. Distribution of ion detected at m/z 1496.75 corresponding to a tryptic peptide of the 71 kDa protein synapsin I
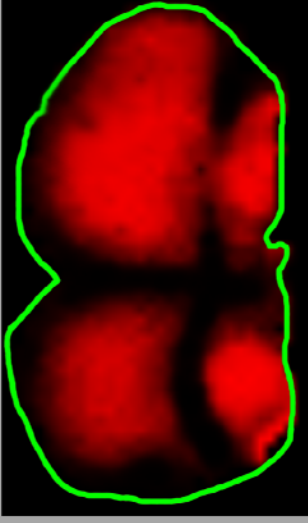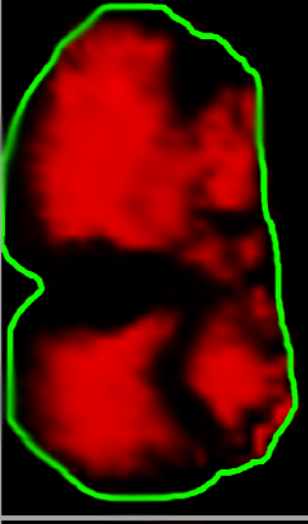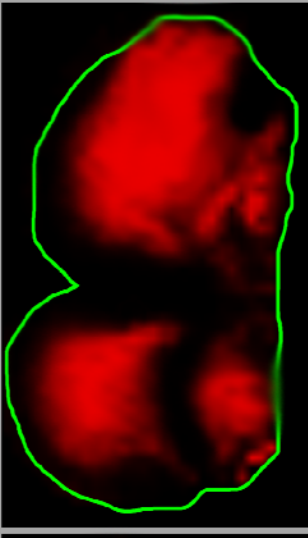
consuming separation and labeling steps commonly used in proteomics with the advantage of preserving the spatial distribution of the detected species.

Technological advancements in instrumentation including high speed electronics, solid state lasers with high repetition rates, enhanced computer processing capabilities, and unique imaging software, continue to make this technology more practical and capable. The development of automated devices for sample preparation [84, 95] has enabled accurate and precise reagent deposition onto microscopic regions of a tissue sample. The ability to digest these protein species and carry out MS/MS analysis directly on tissue allows one to achieve protein identification with high confidence within localized regions of a tissue section without separation or homogenization. The consistent ion intensities within and between separate images indicates that well-controlled and reproducible digestion has been achieved, providing an increased confidence in comparative studies. Digestion also presents possibilities for the detection of certain high molecular weight proteins which are not easily detected in MALDI experiments due to their low abundance and poor ionization and detection efficiency in this mass region.

The identification of proteins directly from tissue combining *in situ* digestion and MALDI MS is most effective in the validation process, i.e., verifying the presence of a protein whose I.D. has either been previously established in other like tissues or is tentatively assigned due to known biology or from other data. The process consists of 5 steps: 1.) establishing the distribution of one or more proteins by molecular weight using imaging MALDI MS, 2.) tryptic

hydrolysis of discrete spots where proteins of interest are located, 3.) acquiring mass spectra from these digested spots using MALDI MS, noting those peptides that correlate with the theoretical peptides that should be generated from a given protein or proteins, 4.) MS/MS sequence analysis and protein database matching for identification of the proteins, and 5.) correlation of the images for each tryptic peptides to that of the intact protein in the tissue specimen.

Although the proteins identified in this work are generally expressed at high abundance, current work towards further optimization of the current protocol is focused on the analysis of lower abundance proteins. For example, the time dependence of on-tissue digestions is an area of much interest that needs further investigation. Ericsson et al.[96] have shown that the downsizing the enzymatic digestion process to nL volumes results in a decrease in the time required to achieve sufficient digestion due to a decrease in the diffusion distance between the enzyme and substrate. A systematic study will need to be performed to determine optimal spotting conditions for digestions that vary enzyme concentration, time, volume, and solvent composition. Other steps such as denaturing of proteins and reducing cysteine-cysteine disulfide bonds prior to digestion could result in more complete digestion. Additionally, other enzymes or combinations of enzymes could be applied to tissues in an effort to carry out more targeted analyses for specific proteins. For example, **Figure 17** shows the distribution of three peptides identified from myelin basic protein in separate rat brain sections using three different proteases: trypsin, chymotrypsin, and subtilisin. For the analysis of complex mixtures such as a protein digest, Fourier

Figure 17. Distribution of three peptides from myelin basic protein generated using three different proteases on separate rat brain tissue sections.

transform ion cyclotron resonance mass spectrometry (FTICR-MS) will be of significant advantage in resolving species in these samples by providing high mass accuracy information.

On-tissue proteomic technologies present enormous potential for the discovery of new protein and peptide species, or modifications of those already known, since it does not require protein-specific reagents such as antibodies. The protocols and technology cited in this paper offer a viable alternative to tissue extraction and separation procedures that nicely complements traditional protein identification and characterization techniques. By combining the extensive amount of data obtained from each of the protein, peptide, and digest imaging experiments, it is possible to carry out an in-depth characterization of a tissue for those proteins amenable to this process. We believe analysis of a tissue in this manner will prove essential in studying disease because it enables the tissue to be examined in its native state.

Materials and Methods

Material. HPLC grade methanol, trifluoroacetic acid (TFA), glacial acetic acid, and reagent grade ethanol were purchased from Fisher Scientific (Pittsburgh, PA). DHB (2, 5-Dihydroxybenzoic acid, 99% purity) was purchased from Acros Organics (New Jersey, USA) and used without further purification. Sinapinic acid (3, 5-dimethoxy-4-hydroxycinnamic acid, 99% purity) was purchased from Fluka (Buchs, Switzerland) and used without further purification. Trypsin Gold was purchased from Promega (Madison, WI) and diluted in 200 μL of 50 mM acetic

acid to obtain a final concentration of 0.5 $\mu g/\mu L$ for the stock solution. A 40 μL aliquot of this stock solution was activated by adding 200 μL of 100 mM ammonium bicarbonate to reach a pH of ~8 and a final trypsin concentration of 0.083 $\mu g/\mu L$. Peptide: N-Glycosidase F (PNGase F) was purchased from New England BioLabs (Ipswich, MA).

Tissue Preparation. Brains and pituitary glands from adult Sprauge-Dawley rats were dissected and stored at -80º C until analysis. Thin (12 μm) tissue sections were prepared and thaw mounted onto a gold plated MALDI target using a procedure described by Schwartz et al.[27]. The targets were placed in a desiccator for 30 min to allow the tissue sections to dry and equilibrate to room temperature. Tissue fixation and removal of salts and other contaminants was carried out through a series of ethanol/water wash steps as described by Aerni et al.[84] An additional final tissue wash step similar to that described by Kutz et al.[76] was used which consisted of submersion of the tissue sections into a solution of 90% ethanol, 9% glacial acetic acid, and 1% deionized water for 30s. The tissue sections were then dried and stored in a vacuum desiccator until analysis.

This tissue preparation procedure removes interfering species, such as salts and phospholipids, that can promote adduct formation, ion suppression, and poor matrix crystallization. Additionally, signals generated from the phospholipids on an un-washed tissue section may result in significant spectral interference with low molecular weight tryptic peptides.

Imaging Mass Spectrometry of Proteins and Peptides. A piezoelectric based chemical inkjet printer (ChIP-1000, Shimadzu Co., Kyoto, Japan) was used for deposition of both the trypsin and matrix solutions. A modified sample holder was used to place gold coated MALDI plates containing tissue sections directly into the robotic printer. Matrix solutions consisting of 15 mg/mL of sinapinic acid in 60:40 ACN:$H_2$0 / 0.5% TFA (aq.) for proteins and 25 mg/mL of DHB in 50:50 MeOH/H_2O / 0.5% TFA (aq.) for peptides were printed onto sections serial to the section used for on-tissue digestion. The arrays of approximately 2500 spots consisted of 250 μm center-to-center spacing and a total matrix volume of ~10 nL at each spot was deposited over a series of 20 iterations at 5 drops per iteration. The images were acquired using an Ultraflex II MALDI-TOF TOF instrument (Bruker-Daltonics Billerica, MA).

For protein imaging, the mass spectrometer was operated with positive polarity in linear mode and spectra were acquired in the range of m/z 1500-30,000. For peptides, the mass spectrometer was operated with positive polarity in reflectron mode and spectra were acquired in the range of m/z 600-10000. A total of 600 spectra were acquired at each spot position at a laser frequency of 200 Hz. Specifically developed software was then used to convert the spectral information into image files compatible with Biomap imaging software (Novartis, Basel, Switzerland) that was used to visualize ion density maps.

On-Tissue Digestion. A solution containing 0.083 μg/μL of trypsin was spotted onto the rat brain tissue sections using the automated printer in an array

incorporating 250 μm center to center spacing between individual spots, each of which were approximately 200 μm in diameter. The trypsin was spotted over a series of 30 iterations while depositing 5 drops (100 pL per drop) each iteration to achieve a total spot volume of 15 nL (5 drops of the trypsin solution were deposited at each position in 8 min. time intervals).   Each digest spot appeared to dry completely between each subsequent spotting iteration.   The trypsin spotting proceeded at room temperature (~21 °C) over a time period of ~4 h, allowing ample time for digestion to take place. Following digestion, a solution consisting of 25 mg/mL of DHB in 1:1 methanol/0.5% TFA (aq). was spotted directly onto the array of tryptic spots over 20 iterations at 5 drops per iteration to give a total matrix spot volume of 10 nL.

Imaging Mass Spectrometry and MS/MS of Digested Tissue Section. The printed arrays were analyzed using a Ultraflex II MALDI-TOF TOF equipped with a smart beam laser[97] and controlled by the Flex Control 2.4 software package. The mass spectrometer was operated with positive polarity in reflectron mode and spectra acquired in the range of m/z 500-6000. A timed ion gate was used for precursor ion selection and the fragments generated were further accelerated with 19 kV in the LIFT cell, and detected following passage through the reflectron. Image acquisition of the spotted arrays was carried out using the Flex Imaging 1.1 software package. A total of 1350 spectra were acquired at each spot position in a customized spiral raster pattern in 50 shot increments at a laser frequency of 200 Hz.   The customized spiral raster pattern was used to

accommodate any edge-biased crystal formation characteristic of DHB matrix spots that may have occurred. Ion images were created as described above.

Glycopeptide Analysis. The posterior region of a 12 μm section of a rat pituitary gland was spotted 3 consecutive times with 500 nL of a 0.083 μg/μL of trypsin using a micropipette, with each drop being allowed to dry before spotting the next. Next, 500 nL of the amidase PNGase F was spotted 3 consecutive times at a concentration of 500,000 U/mL directly onto the trypsin spots, with each spot being allowed to dry before placing the next. Upon drying, 2 drops of 500 nL DHB in 1:1 methanol/0.5% TFA (aq). were deposited at these positions and the tissues were analyzed using a Ultraflex II MALDI-TOF TOF.

Data Analysis. The MS/MS spectra generated were submitted into a MASCOT (Matrix Science, Boston MA) database search engine to match tryptic peptide sequences to their respective intact proteins. The MS/MS spectrum search was performed with a peptide tolerance of ± 0.3 Da and a fragment tolerance of ± 0.5 Da. The search criteria also included up to 2 missed cleavages and variable modifications including lysine acetylation, N-terminus acetylation, C-terminus amidation, and methionine oxidation.

CHAPTER III

HIGH-THROUGHPUT PROTEOMIC ANALYSIS OF FORMALIN-FIXED PARAFFIN-EMBEDDED TISSUE MICROARRAYS USING MALDI IMAGING MASS SPECTROMETRY

Abstract

A novel method for high-throughput proteomic analysis of formalin-fixed paraffin-embedded (FFPE) tissue microarrays (TMA) is described using on-tissue tryptic digestion followed by MALDI imaging mass spectrometry. A TMA section containing needle core biopsies from lung tumor patients was analyzed using mass spectrometry and the data correlated to a serial H&E stained section having various histological regions marked including cancer, non-cancer, and normal. By associating each mass spectrum with a defined histological region, statistical classification models were generated that can sufficiently distinguish adenocarcinoma biopsies from squamous cell carcinoma biopsies. These classification models were built using a training set of biopsies in the TMA and were then validated on the remaining biopsies. Peptide markers of interest were identified directly from the TMA section using MALDI MS/MS sequence analysis. The ability to detect and characterize tumor marker proteins for a large cohort of FFPE samples in a high throughput approach will be of significant benefit not only to investigators studying tumor biology, but also to clinicians for diagnostic and prognostic purposes.

Introduction

IMS has been successfully used to identify subsets of markers that correlate with cancer progression.[65, 67, 69, 98-100] For example, in earlier studies, protein markers were obtained from the direct MALDI IMS proteomic analysis of glioma biopsies that could differentiate stage including the most aggressive form, glioblastoma multiforms, and also with patient outcomes.[69] Our laboratory has also investigated a large cohort of human non-small cell lung cancer biopsies.[65, 68] In this study, sections from fresh frozen biopsies were cut and spotted with matrix on areas identified as cancerous. From the resulting protein profiles, statistical analyses identified markers that could be correlated with histological assessment and patient outcomes. These patterns precisely distinguished healthy versus cancerous tissue, and also distinguished various sub types of non-small cell cancer, such as adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Further, these patterns could be correlated with patient survival. From this latter cohort, several of the proteins of the survival signature were identified. These results highlight the use of this technology for the rapid characterization of disease at the protein level to confirm diagnosis and potentially aid in therapeutic management.[101]

MALDI IMS is usually carried out on fresh frozen tissue because these samples are highly representative of a tissue in its native state. However, the vast majority of clinical specimens stored in hospital tissue banks are formalin-fixed paraffin embedded (FFPE), representing an expansive archive of diseased tissues. Formalin fixation stabilizes proteins by inducing chemical cross-linking

throughout a tissue section therefore preventing postmortem enzymatic proteolysis while maintaining the cellular histology.[102]   However, the major drawback of formalin fixation is that it renders a tissue unsuitable to undergo routine biochemical protein extraction and subsequent proteomic analysis.[103]

Over the past two decades, several methods have been reported that attempt to reverse formalin-fixation, a process commonly referred to as antigen retrieval.[104-106] Antigen retrieval typically involves the application of high temperature treatment along with the use of a buffer solution, in an effort to reverse the protein cross-linking and return a tissue to its native state.[104] Development of methods to effectively and reproducibly carry out antigen retrieval on FFPE tissue specimens has allowed for the standardization of IHC protocols and has recently opened the door to a vast collection of archival clinical samples to be analyzed for genomic and proteomic information. For protein identification, approaches using either liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) [106-109] or 2D gel electrophoresis[110] have been reported. Recently, a 2D gel based study showed that the same proteins can be identified independently of the type of preservation used.[111]   These authors reported that when the appropriate antigen retrieval protocol is applied, the level of protein identification was found comparable to that of fresh frozen tissues.

In order to comprehensively evaluate the diagnostic, prognostic, and therapeutic value of gene and/or protein expression in clinical tissue specimens, it is necessary to analyze a large number of samples from patients in different stages of disease.[112] In this regard, tissue microarrays (TMA) were originally

developed to facilitate the molecular and pathological analysis of tissues in a massively parallel and high-throughput approach.[112-114] A TMA consists of a paraffin block in which as many as 1000 cylindrical tissue biopsies from individual tumors are distributed into a precise array.[112] Sections cut from this array enable investigation of DNA by fluorescence *in situ* hybridization (FISH), RNA by mRNA *in situ* hybridization, or proteins by IHC, from each of the biopsy samples.

We have developed an IMS method incorporating antigen retrieval and *in situ* enzymatic digestion to analyze the protein content of tissue microarrays (TMA) containing FFPE non-small-cell lung cancer (NSCLC) biopsies.[115] The TMAs used in this study contained various types of non-small-cell lung cancer (NSCLC) biopsies including squamous cell carcinoma, adenocarcinoma, and bronchioloalveolar carcinoma as well as non-cancer tissue from matched individuals. MALDI IMS was used to analyze these samples in a high-throughput fashion providing proteomic data from discrete regions of each biopsy. We demonstrate, as previously observed at the protein level from fresh frozen tissue [65], that a histological classification of lung cancer can be effectively accomplished at the peptide level on FFPE tissues. The ability to detect and characterize tumor marker proteins for a large cohort of samples in a high throughput approach will be of significant benefit not only to investigators studying tumor biology, but also to clinicians for diagnostic and prognostic purposes.

Results

The antigen retrieval protocol used in these studies enables the analysis of FFPE tissues using the *in situ* digest methodology. **Figure 18** shows the spectra generated on two serial FFPE TMA sections where one section was subjected to the antigen retrieval protocol and the other was not (control). It is clear from these results that antigen retrieval is necessary and effective in allowing the enzyme access to the proteins in these tissues. The spectra generated from the section that underwent antigen retrieval shows hundreds of tryptic peptide peaks while only trypsin autolysis peaks were detected from the control section. Autolysis refers to the process where the trypsin enzyme begins to digest other trypsin molecules. A small degree of autolysis is expected, as in Figure 18 b.), however the excessive amount of autolysis shown in Figure 18 a.) is indicative of virtually no proteins being accessible to the enzyme.

The FFPE TMA analyzed in this study contains 1 mm needle core biopsies from lung tumors diagnosed as adenocarcinoma and squamous cell carcinoma, as well as adjacent normal control tissue and other NSCLC tumors. This study only considers the adenocarcinoma and squamous cell carcinoma biopsies due to the limited number of samples for the other cancer types. The layout of the TMA analyzed is such that there are duplicate and, in some cases, triplicate needle core punches for several of the biopsies. Therefore, the TMA contains a total of 24 squamous cell carcinoma needle cores from 15 different patients (7 unpaired, 7 duplicates, and 1 triplicate) and 19 adenocarcinoma
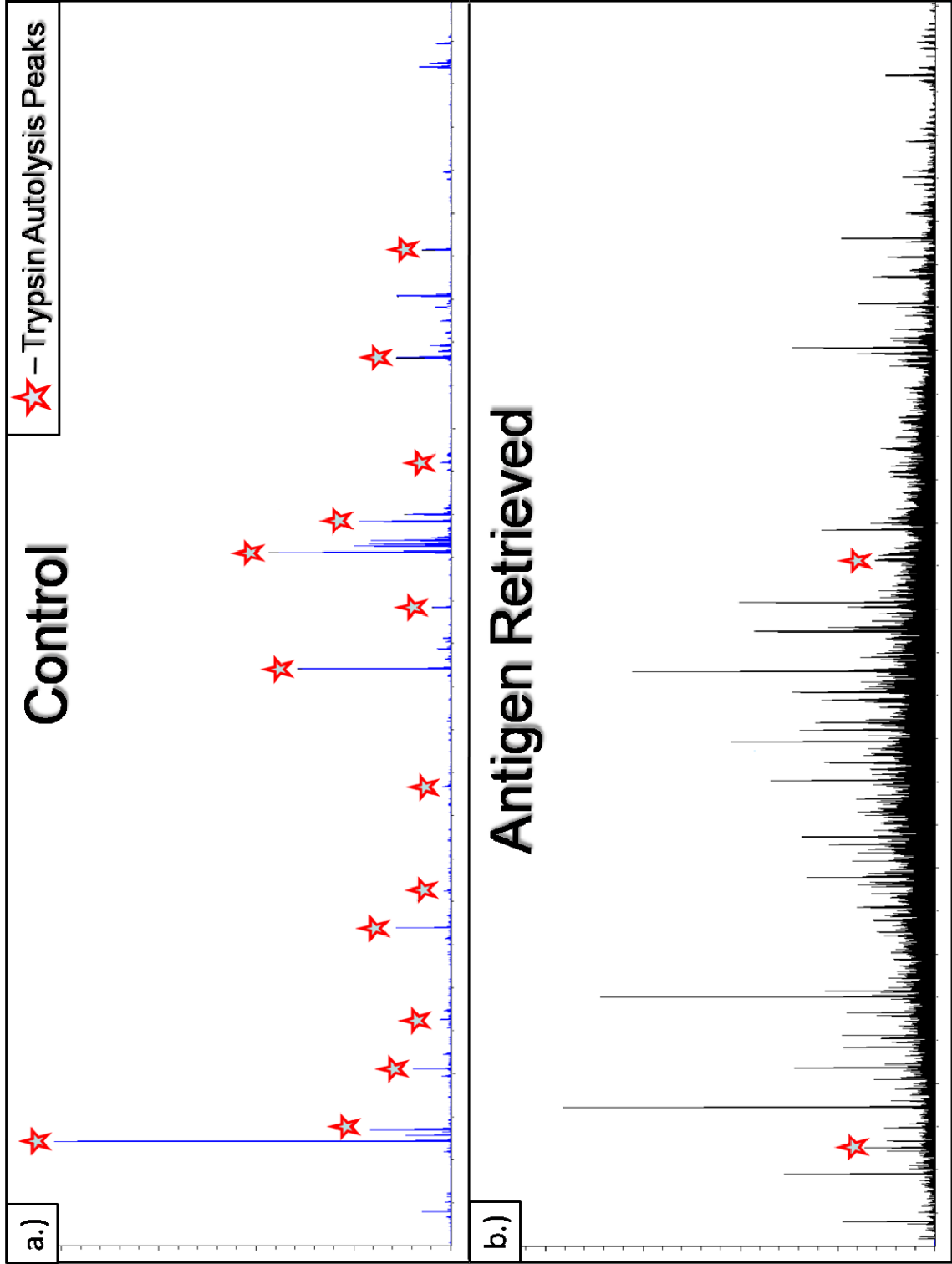
Figure 18 - a) TMA control section analyzed using *in situ* digestion without antigen retrieval **b)** TMA section analyzed using *in situ* digestion with antigen retrieval.

needle cores from 11 different patients (3 unpaired and 8 duplicates) that contain cancerous regions and are considered in the study.

In the discovery and protocol development process, a hematoxylin and eosin (H&E) stained section of the TMA was first analyzed using light microscopy and the cancer, non-cancer, and normal regions were marked in each biopsy based on histology. Non-cancer regions consist of tissue that is not cancerous but is not normal either, including areas of inflammation, scar tissue, etc. The cellular regions that were not able to be clearly identified were not marked and therefore were not used in the statistical model generation step. This annotated tissue section was then co-registered with a serial TMA section analyzed by MALDI IMS, (**Figure 19**), enabling individual mass spectra obtained from each coordinate position on the tissue (pixel) to be linked to that same precise histological region in the TMA.

The mass spectrum generated at each spot on the digested tissue typically contain many hundreds of peaks with a signal-to-noise (S/N) > 5 (**Figure 20 (a)**). The ensemble of tryptic peptides at each position represents a variety of proteins with a broad range of functionality and molecular weights. The signal intensities of the tryptic peptides are mediated by several factors, i.e. protein concentration differences, variations in enzymatic digestion efficiency, and differences in desorption and ionization efficiencies. Nonetheless, mass spectra from similar histological regions contain peak profiles with a high degree of concordance, demonstrating a consistent and reproducible methodology as shown in **Figure 21**. This figure illustrates, as expected, that the average spectra
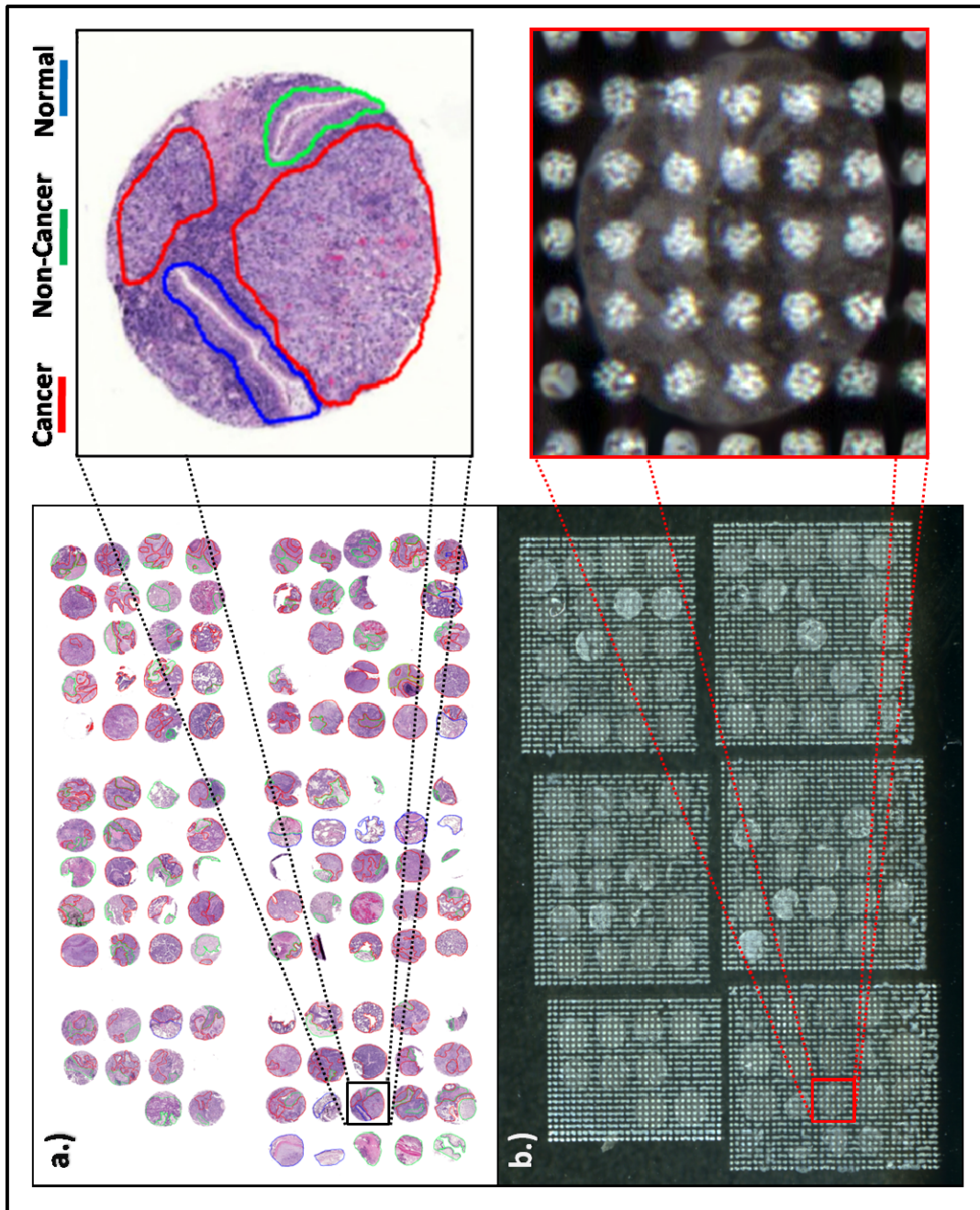
Figure 19 - a) TMA H&E with histological regions marked **b)** TMA spotted with trypsin/matrix for MS analysis
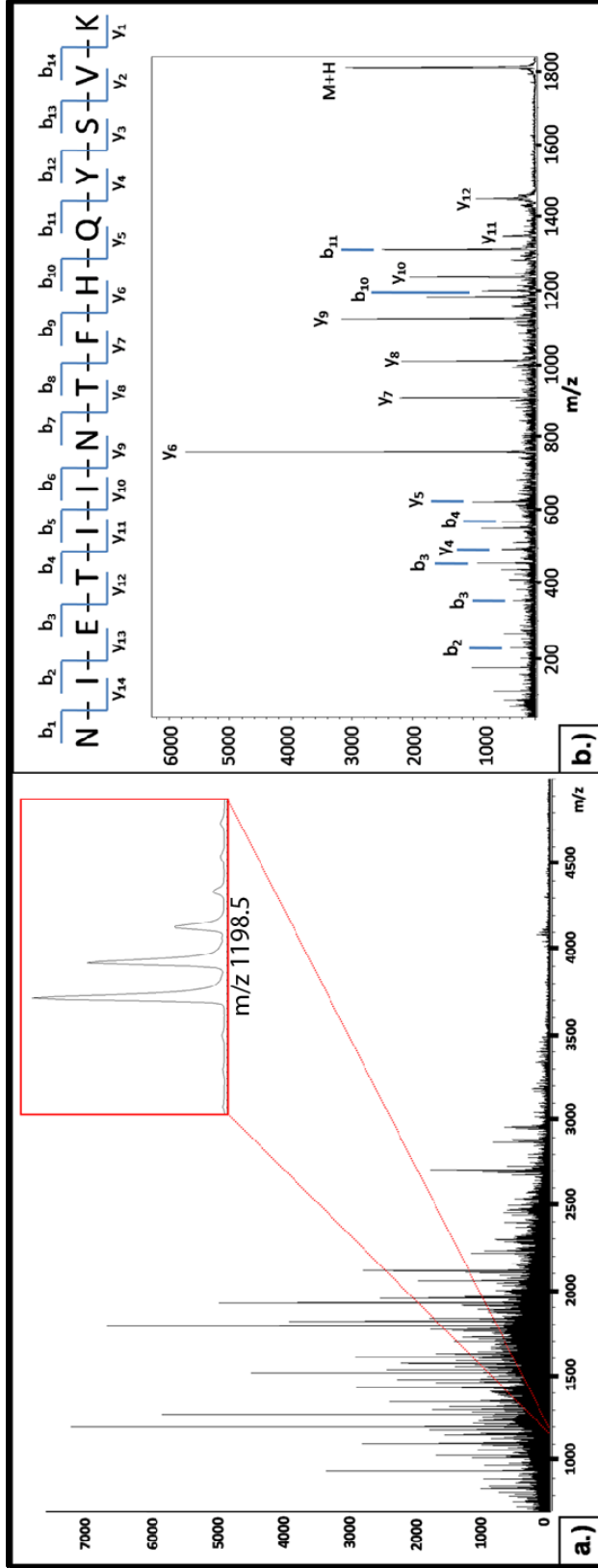
Figure 20 a) Representative spectrum acquired from an individual digest spot in a cancerous region of an adenocarcinoma biopsy **b)** Example of an MS/MS spectrum acquired directly from the FFPE TMA and sequenced as a tryptic peptide from the protein S100-A9
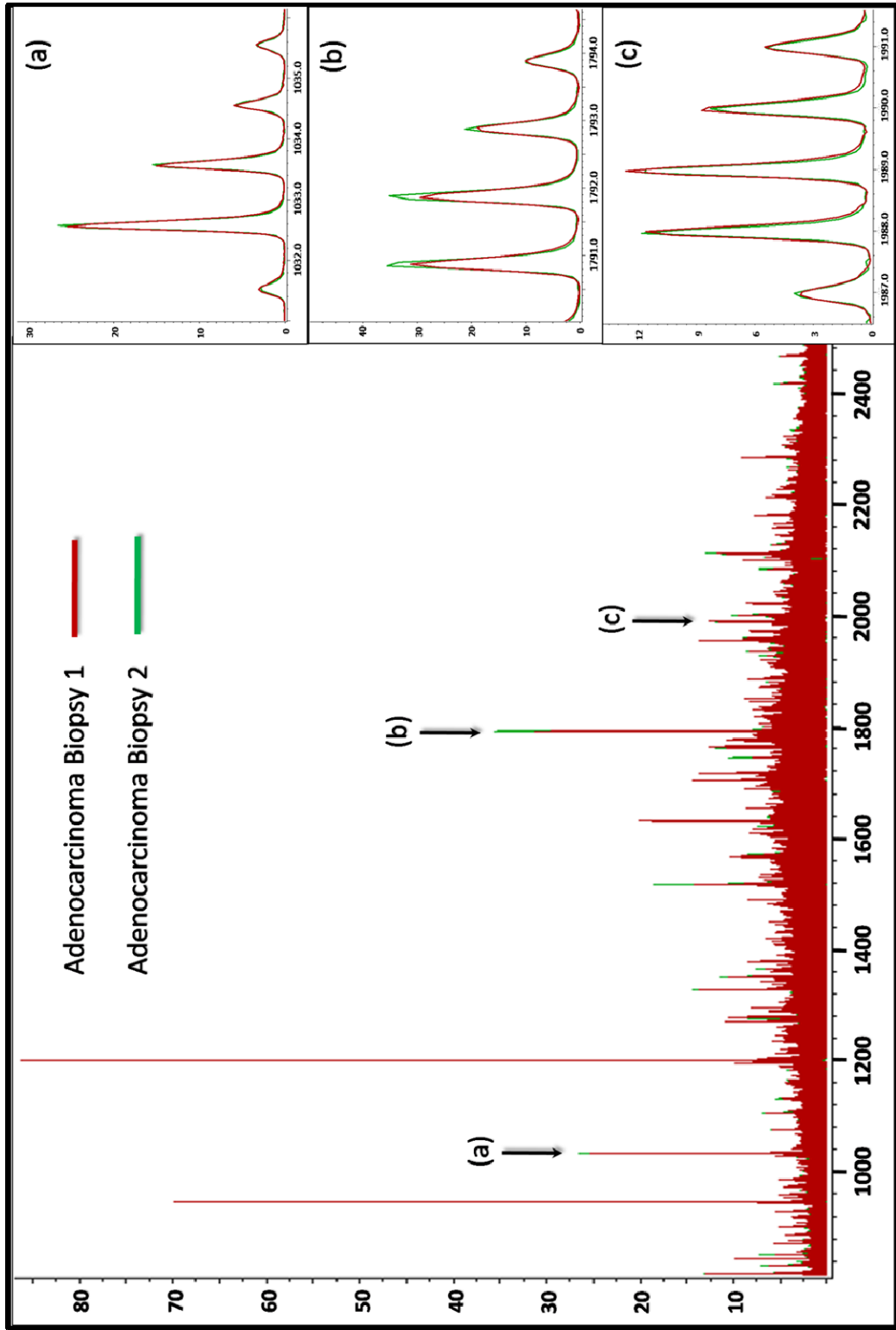
Figure 21 - Overlay of the average spectrum for two duplicate needle cores in the TMA from the same adenocarcinoma biopsy

generated from the cancer regions of two duplicate biopsies are extremely similar. This reproducibility is further exhibited using a t-test and minimum 1.6-fold intensity difference comparison for the 200 most intense peaks in the average spectrum from the cancer regions of each biopsy. Based on these criteria, it was determined that there were no statistically significant differences in any of the peak intensities between the two biopsies. It should be noted that these are two different needle cores form the same biopsy and thus are not identical pieces of tissue.

The protease hydrolysis step is essential to generate peptide fragments derived from non-crosslinked domains and enable identification directly from their location in the tissue. Typically, the process generates hundreds of tryptic peptides in a mass range (*m/z* 500-3000) amenable for sequence analysis using a MALDI-TOF/TOF instrument. Thus, selected peptides are desorbed and sequenced directly from the tissue and subsequently linked to the respective intact proteins originally present.

Currently, we have identified ~50 proteins directly from the lung tumor TMA using MALDI MS/MS sequence analysis (**Appendix A**). An example of an MS/MS spectrum acquired directly from the FFPE-TMA and sequenced as a peptide from the protein S100-A9 is shown in **Figure 20 (b).** Additionally, five fully annotated MS/MS spectra are included in the **Appendix B** to provide a detailed illustration of the spectral quality that can be obtained from carrying out MS/MS with a MALDI-TOF/TOF instrument directly from a tissue.

The ability to visualize the spatial localization of a set of tryptic peptides generated from a single parent protein provides an additional level of validation, given that the distribution of these species should be identical. For example, heat shock protein beta-1 was found to be localized almost exclusively to the cancer regions of a subset of the squamous cell carcinoma biopsies. Three tryptic peptides from this protein were detected and sequenced using MS/MS sequence analysis, including m/z 987.60, 1163.62, and 1905.99. **Figure 22(a)** shows the average spectra for three different squamous cell carcinoma biopsies, including two from which the tryptic peptides for heat shock protein beta-1 were was detected and identified, and using MS/MS sequence analysis, as well as a third biopsy where this protein was not detected. The zoomed spectra view for the three tryptic peptides from this heat shock protein beta-1 show the consistent peak distributions for proteins detected at similar levels in different tissue samples. Furthermore, **Figure 22(b)** illustrates the similar ion density distribution of these three peptides across the TMA.

The layout of the TMA used in these analyses is shown in **Figure 23(a)**. This TMA was constructed to contain duplicate needle cores from several of the tumor biopsies along with unpaired biopsies and a set of biopsies from adjacent non-involved normal lung tissue. For statistical analysis, the duplicate biopsies are separated into two datasets; set 1 is used as a training set to build the classification models and set 2 is used to evaluate the accuracy of this model through cross-validation. The first step in analyzing the data set generated from the imaging experiment is to develop a classification model that can
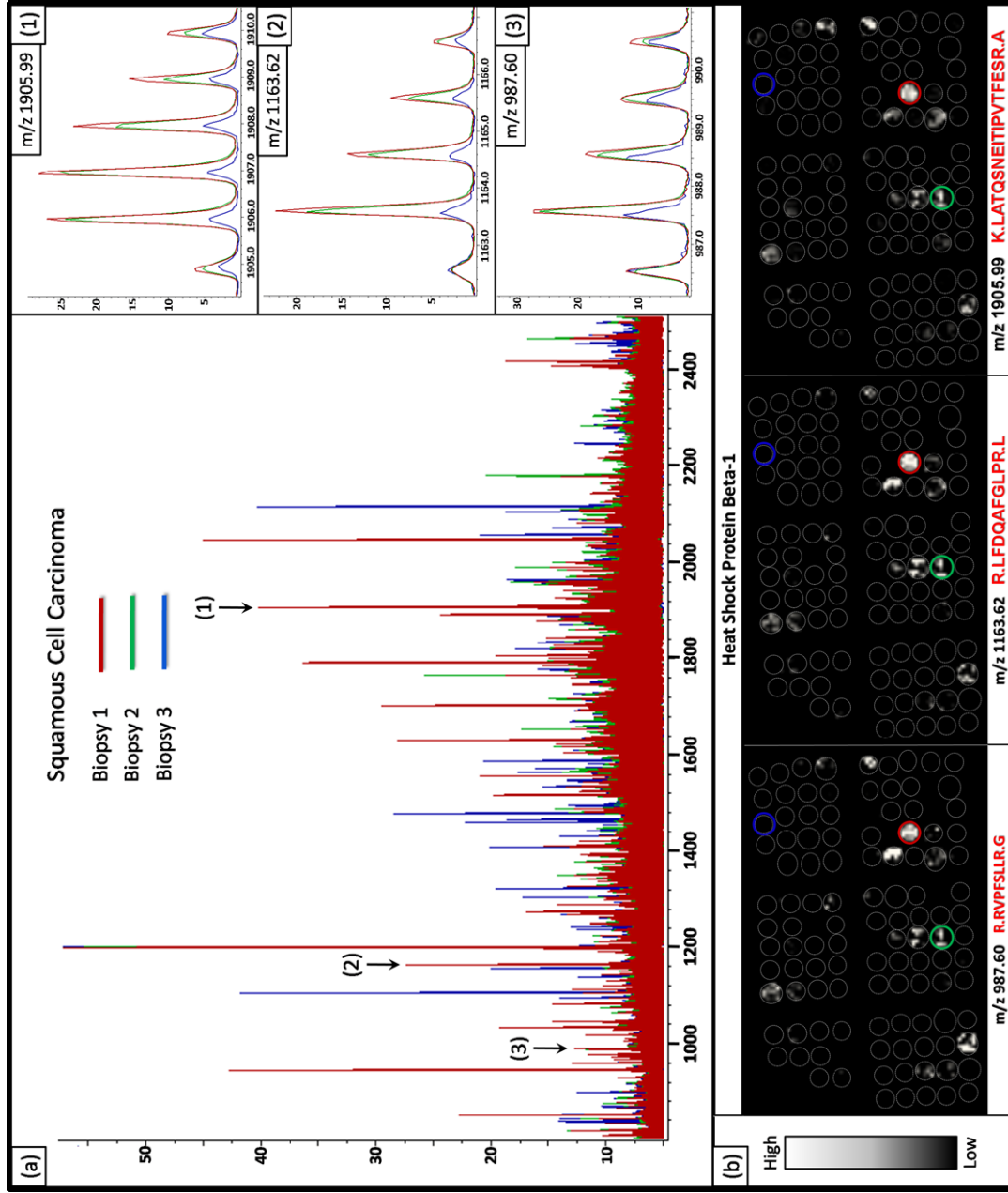
Figure 22 - a) Overlay of average spectrum from three squamous cell carcinoma biopsies taken from separate patients **b)** The distribution of three tryptic peptides from heat shock protein beta-1
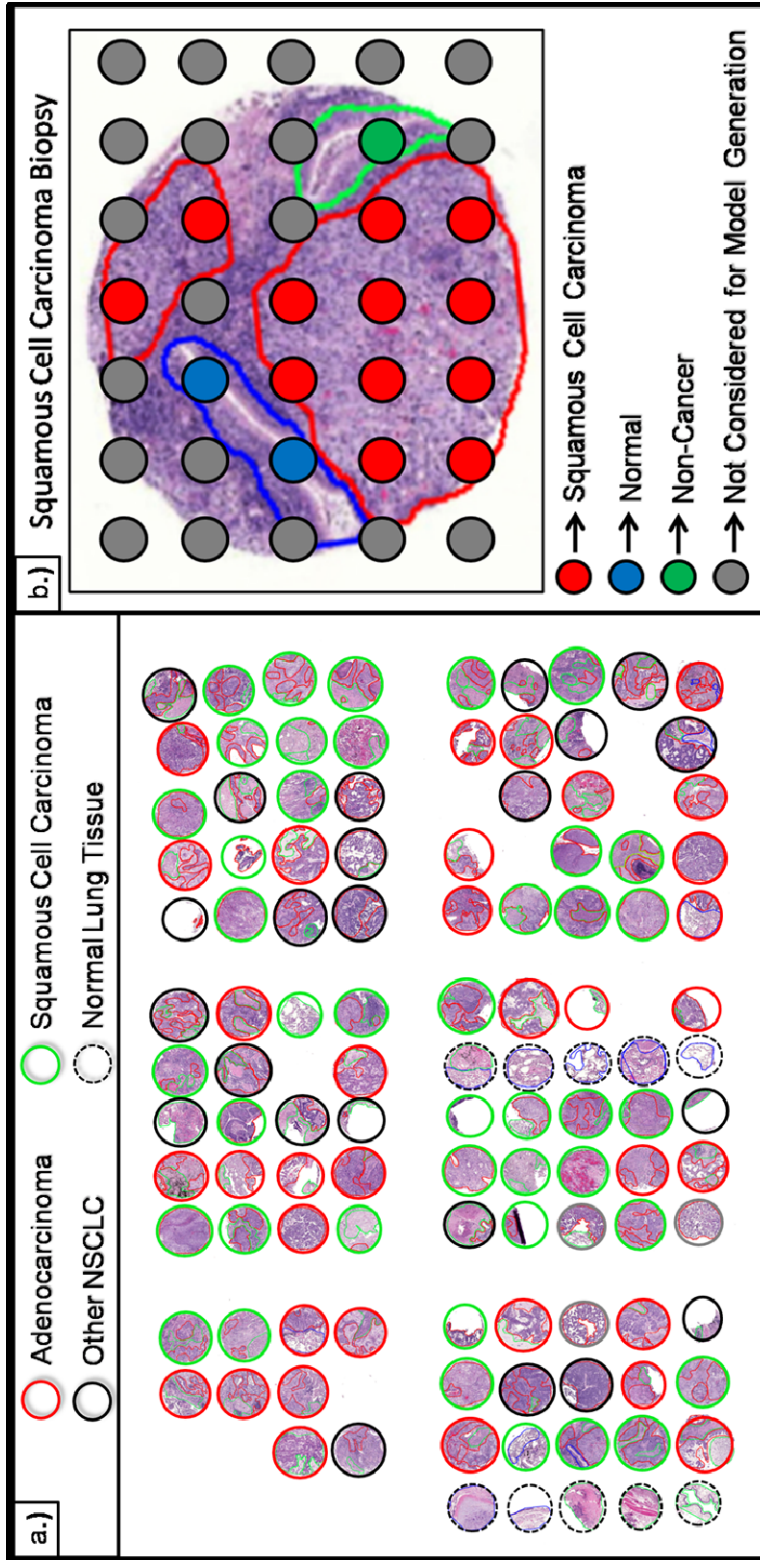
Figure 23 - a) H&E stain of a TMA section with pathological diagnosis outlining each biopsy **b)** Each individual spectra is grouped based on the histological region from which it was acquired

differentiate the different cancer regions from the non-cancer and normal regions. The classification models are generated by first grouping together the spectra from each of the cancer and non-cancer regions into separate folders for each biopsy in the TMA. This is completed using the cancer diagnoses and marked histological regions on each biopsy as shown in **Figure 23(b).** Since the average size of a matrix spot in these analyses is ~175 μm, it is possible that some matrix spots are positioned across multiple histological regions, creating a mass spectrum partially representative of each of the underlying cell types. In the training phase, only spectra that are clearly located within a single histological region are included for generating the statistical classification models. The spectra which meet this positional criterion in set 1 of the TMA patient biopsies are exported into the appropriate adenocarcinoma, squamous cell carcinoma, non-cancer, or normal class groupings to be used for model generation. For example, **Figure 24** shows the average spectra for a squamous cell carcinoma needle core biopsy overlaid with the average spectra from a needle core biopsy taken from adjacent normal tissue from same patient. The peak distributions in these two spectra are very different and there are clearly a large number of peptides present that can be used as class identifiers. Several approaches can be used to systematically identify peaks that sufficiently distinguish the various tumor classes from each other. The most straightforward way to do this is to combine the spectra in each group to create an average spectrum representative of a class identified through histology. The peaks present in these average spectra can then be compared through statistical analysis to identify a subset of

70
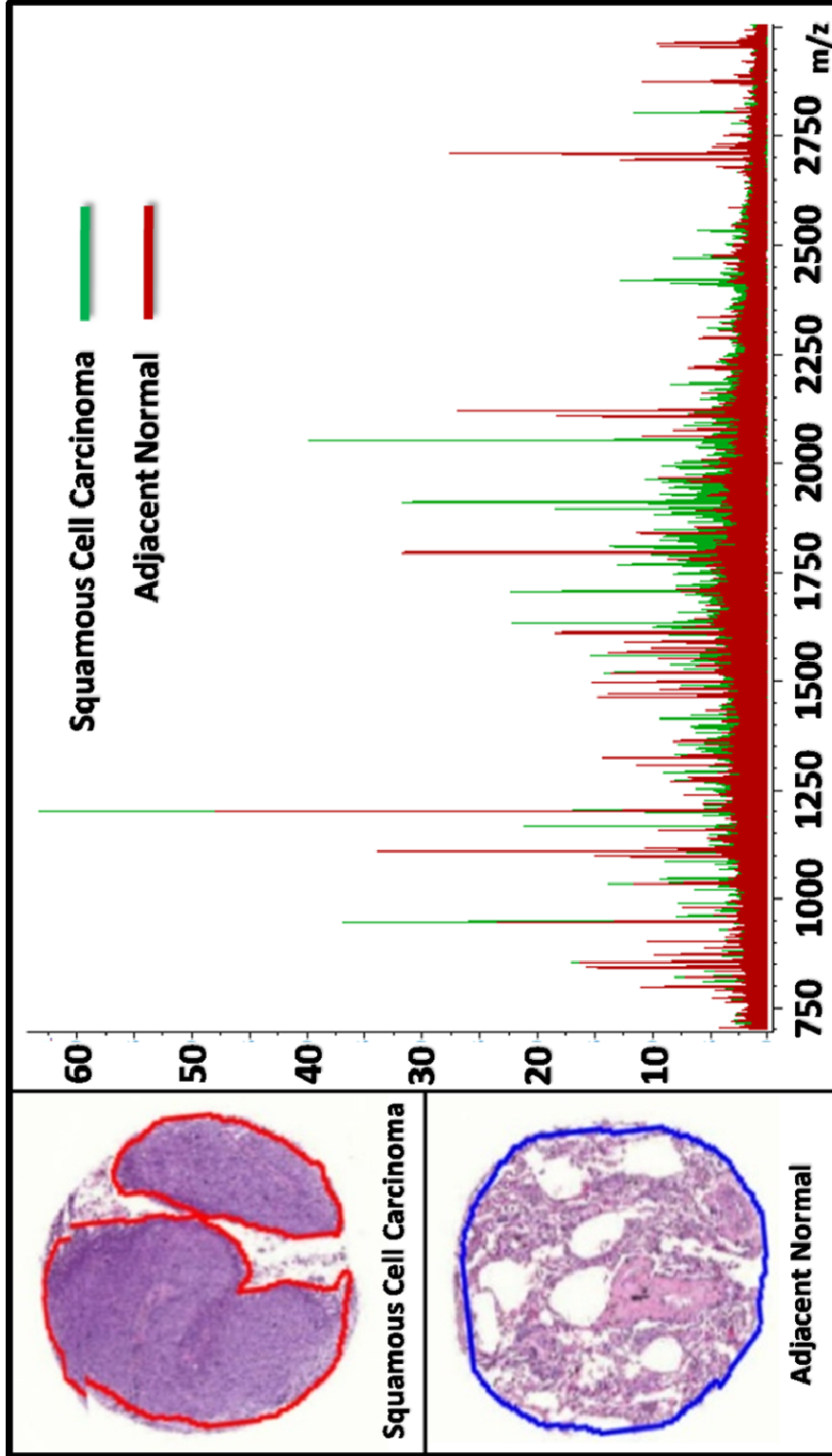
Figure 24- Overlay of average spectra from a squamous cell carcinoma needle core biopsy with the average spectra from a needle core biopsy taken from adjacent normal tissue

peaks that are significantly different between each group and can therefore be used as the class identifiers. To do this, the statistical software (ClinPro Tools, Bruker Daltonics) was used to create a table of average intensity and standard deviations for a set of the 200 most intense peaks in the average spectrum for both the adenocarcinoma and squamous cell carcinoma samples. The set of 200 peaks were evaluated between samples using a t-test and minimum three-fold intensity difference average comparison to determine a list of class identifiers. The limitation of this method is that it is possible that peaks which may be significant among a small subset of spectra in a group will become insignificant when averaged with the other spectra in that group. For example, the peptide at m/z 1410.7, a tryptic peptide from keratin, type II cytoskeletal 5 (KRT5), shows a distribution localized exclusively to a subset of the squamous cell carcinoma biopsies (**Figure 25**). This is further illustrated when a statistical comparison of the average peak intensity in the squamous cell carcinoma biopsies for m/z 1410.7 is compared to the set of adenocarcinoma biopsies. A significance analysis using a minimum two-fold intensity difference threshold and a t-test shows that when all spectra from each of the adenocarcinoma and squamous cell carcinoma biopsies in the training set are combined into two average spectra, m/z 1410.7 is not a significant class identifier. However, when a subgroup of squamous cell carcinoma biopsies which do express cytokeratin 5 are averaged and compared to the adenocarcinoma biopsies, this peak is a significant classifier.

72
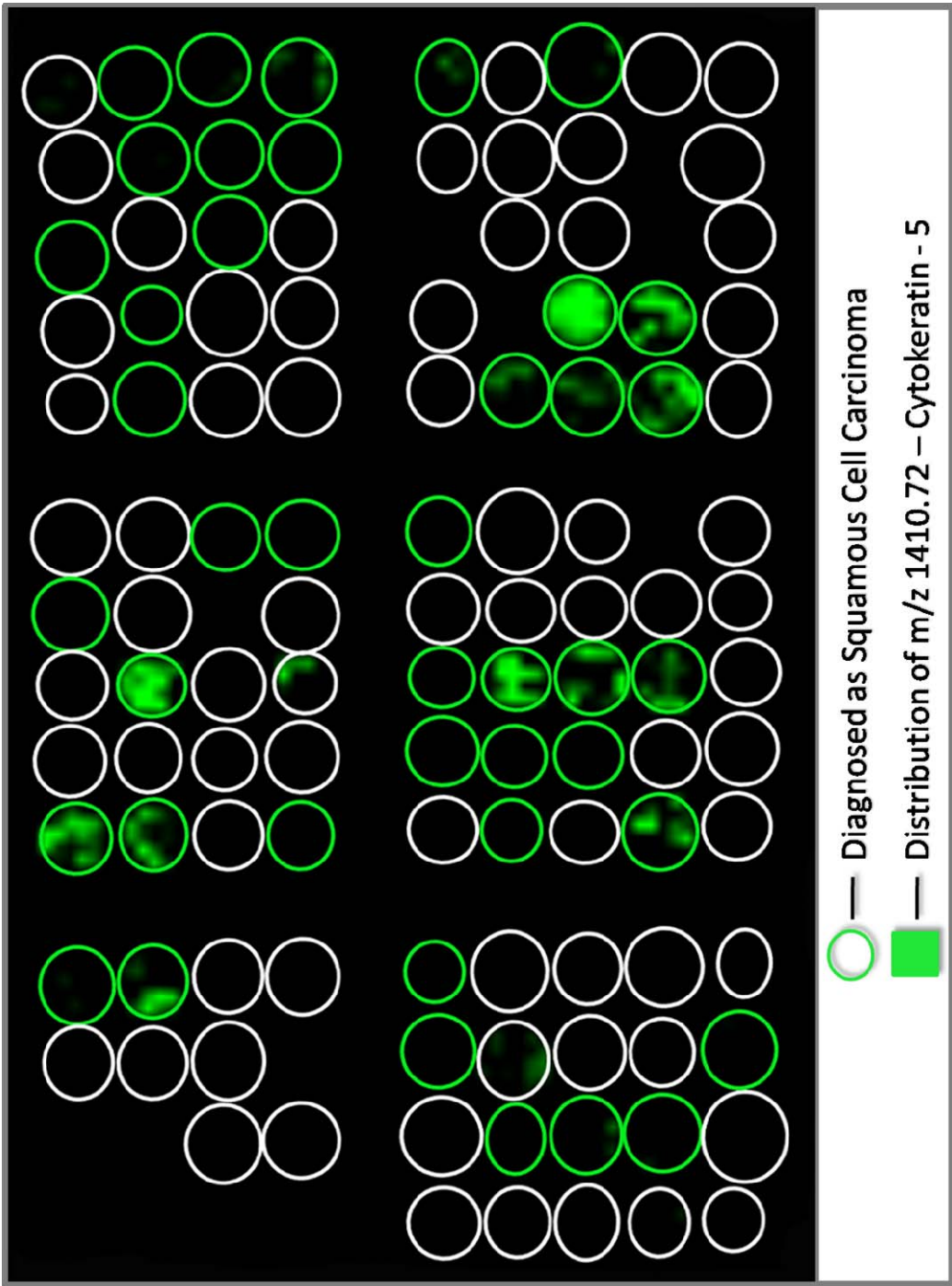
Figure 25 - Ion density map of m/z 1410.70 showing an exclusive distribution to a subset of the squamous cell carcinoma biopsies

Legend:
— Diagnosed as Squamous Cell Carcinoma
— Distribution of m/z 1410.72 — Cytokeratin - 5

To address the heterogeneity of peak distributions for biopsies with the same diagnosis, it may be necessary to first cluster the samples within each histology class based on a similarity criteria. This is done by comparing the average intensity and standard deviation calculated in the average spectrum for each individual adenocarcinoma sample in the TMA to the values in the average spectrum for each individual squamous cell carcinoma sample and vice versa. A similarity value for each comparison was calculated by assessing the number of peaks out of the 200 that were determined to be significantly different between the two samples. This value is used to evaluate the heterogeneity of peak distributions within a class (i.e. adenocarcinoma) and between classes (i.e. adenocarcinoma vs. squamous cell carcinoma). This similarity value is used as a general measure of the reproducibility for this method by comparing the peak distributions of two different needle core biopsies that were taken from the same patient sample. A list of class identifiers was developed by tabulating the frequency that each peak is determined to be statistically different between the individual adenocarcinoma and squamous cell carcinoma samples. These peaks are used to cluster patient samples with the same pathological diagnosis into subclasses with other samples having a similar peak distribution.

It was also possible to classify the adenocarcinoma and squamous cell carcinoma regions successfully using the combined average spectra for each group to develop a peak list of class identifiers. The classification models were built using a support vector machine (SVM) algorithm which can then classify spectra based on the supervised learning from the training set. The SVM

algorithm used in these experiments incorporated 73 peaks, determined through statistical comparison of the peaks in the training sets for both adenocarcinoma and squamous cell carcinoma as described above. The SVM model was run against all spectra in the data set and the outcome of each classification was visualized using the class imaging function in FlexImaging 2.0. For example, **Figure 26** displays the statistical classification of four TMA biopsies that are in agreement with the diagnosis made based on histology. This model classified the spectra from regions marked as adenocarcinoma with an accuracy of 97.9% (140/143 spectra) and squamous cell carcinoma with an accuracy of 98.6% (141/143 spectra). Another way to evaluate these data is a comparison of individual patients and on this basis, all 19 of the adenocarcinoma biopsies and all 24 of the squamous cell carcinoma biopsies were classified correctly. In the adenocarcinoma, the 3 misclassified spectra were dispersed randomly throughout the dataset as single occurrences and therefore had little effect on the overall patient classifications. In the squamous cell carcinoma dataset, the two misclassified spectra came from a single biopsy in which the remaining 9 out of the total 11 spectra were classified correctly as squamous cell carcinoma. It is noted that the MS data was compared to normal histology/pathology, itself not a gold standard.
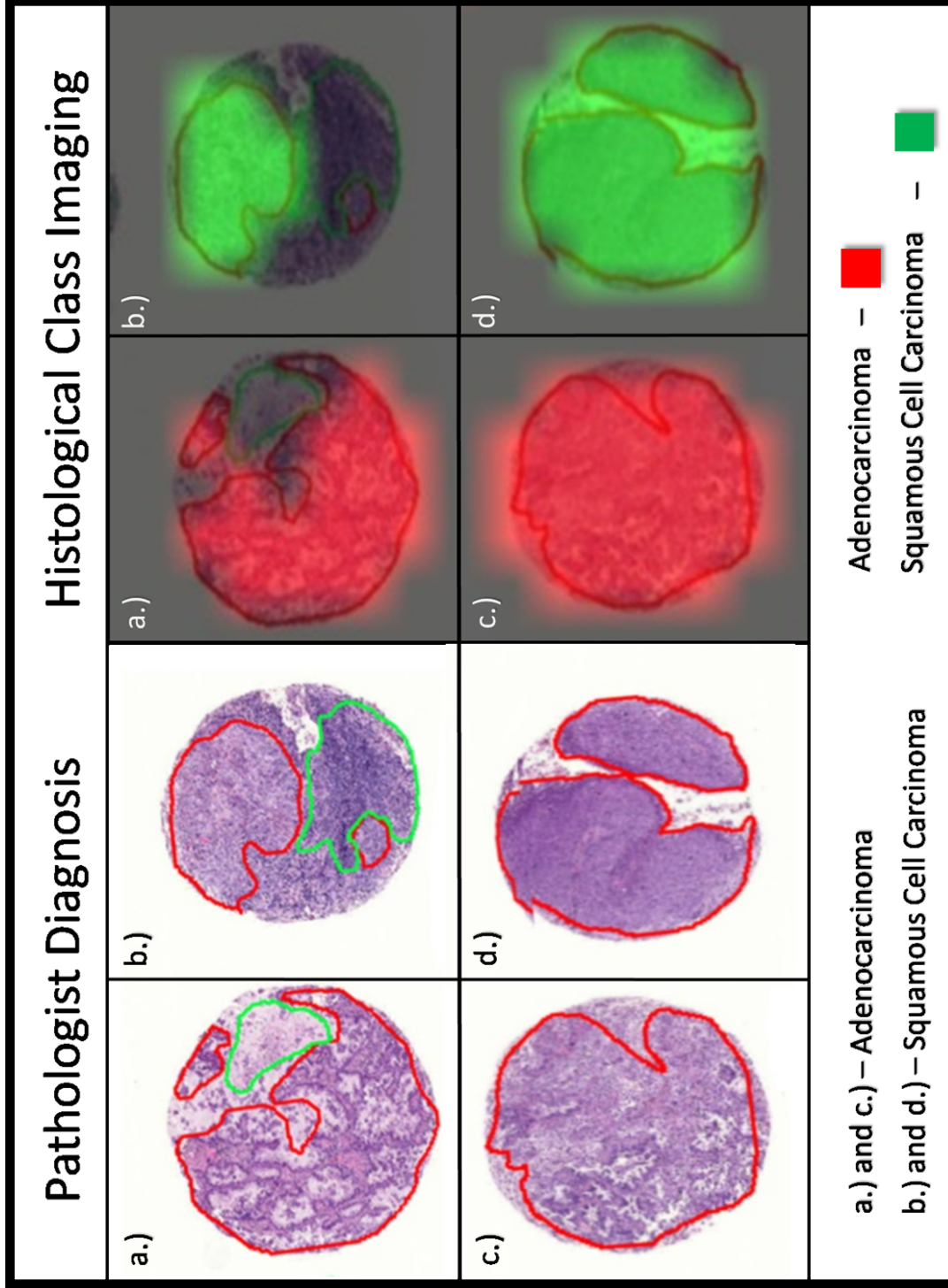
Figure 26. Visual representation of the statistical classification of 4 biopsies compared to the marking and diagnosis of the pathologist

Discussion

A comprehensive understanding of the mechanisms behind carcinogenesis, tumor progression, and metastasis will require an in-depth analysis of not only the genome, but also the proteome, that direct central components in the signaling pathways that drive neoplasia.[116] Whereas mutations in the genomic contents of cells are known to be the principal cause of the onset of carcinogenesis, detecting the changes induced at the protein expression level should prove invaluable in characterizing the molecular machinery responsible for neoplasia.[116] These protein patterns should provide information of the underlying cellular processes that control the development of cancer and subsequent metastasis. Although mRNA patterns can be used to define subclasses and prognostic subsets of lung carcinomas [65, 117-119], they generally are not able to identify the proteins expressed in a tissue and how these proteins are modulated. Therefore, direct analysis of the proteome in cancer tissues may provide a more accurate representation of the current pathological state.

In this work we provide evidence that using on-tissue digestion coupled with imaging mass spectrometry (IMS) and statistical classification distinguishes different lung cancer histology's and sub-classify individual cancer types. Protein expression and relative quantification data can be generated for multiple patient tissue samples in a single experiment. This creates a platform for comparison which could be valuable in determining protein markers indicative of various

disease states and other clinical information such as prognosis and treatment effectiveness.

The data presented is based on a single TMA for a small patient cohort. The experiments described in this work were aimed at demonstrating the feasibility of classification of tumor samples in a TMA at the molecular level using MALDI IMS. In order to carry out a more systematic biomarker discovery, it will be necessary to evaluate the classification models built using this sample group on a much larger set (>100) of patient biopsies. This will also be essential in determining the true diagnostic and prognostic value this method of tissue analysis can offer. In order to effectively manage the enormous dataset generated in such experiments, new bioinformatics tools and workflows will need to be developed that better allow the user to evaluate model generation and classification process.

One of the primary advantages of IMS is the visualization of the molecular content of a sample while maintaining the spatial integrity. This aspect proves essential for the analysis of samples which contain a heterogeneous distribution of cell types. For example, resected lung tumor biopsies exhibit a wide range of cellular morphologies, often containing areas of cancer cells dispersed within normal and preneoplastic tissue regions. Analysis of the proteome from these various regions requires that the cellular structure of the tissue be maintained. *In situ* trypsin digestion provides this distinct capability and, when coupled with IMS, provides a descriptive analysis of a tissue's protein contents relatively quickly in a single experiment. Protein identification directly from tissue is not always

straightforward because of the large number of peptides generated from proteolysis. We have recently begun analyzing these samples using a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS) in an effort to address the complexity of the spectra through high resolution and high mass accuracy detection. Although the time-of-flight (TOF) mass analyzer has extraordinary analysis speed and throughput, it lacks the mass resolution needed to resolve all off the detected ions. Nonetheless, the advantage of maintaining the spatial location of the identified species is critical in achieving a comprehensive analysis of these highly heterogeneous tissue samples.

Materials and Methods

Materials. HPLC grade acetonitrile (ACN) and trifluoroacetic acid (TFA) were purchased from Fisher Scientific (Pittsburgh, PA). Alpha-cyano-4-hydroxy-cinnamic acid (CHCA) was purchased from Fluka (Buchs, Switzerland) and used without further purification. Trypsin Gold was purchased from Promega (Madison, WI). ITO-coated conductive slides were purchased from Delta-Technologies (Stillwater, MN). Tissue microarray (TMA) consisting of 100 duplicate needle core biopsies (1 mm) from 50 patients diagnosed with non-small cell lung cancer and 10 adjacent normal lung tissue punches. The layout of the TMA includes duplicate needle core punches from 21 adenocarcinoma biopsies, 21 squamous cell carcinoma biopsies, 4 bronchioloalveolar carcinoma biopsies, 2 metastatic colon cancer biopsies, 1 carcinoid biopsy, and 1 plasma cell granuloma biopsy. Due to limited numbers for the other cancer types only the adenocarcinoma and

squamous cell carcinoma biopsies are considered in these experiments. Several of the needle cores from the adenocarcinoma and squamous cell carcinoma biopsies did not contain a sufficient amount of tissue or did not contain regions of cancer. Therefore, the TMA contains a total of 24 squamous cell carcinoma needle cores from 15 different patients (7 unpaired, 7 duplicates, and 1 triplicate) and 19 adenocarcinoma needle cores from 11 different patients (7 unpaired and 7 duplicates) that contain cancerous regions and are considered in the study.

Tissue Preparation. Serial 5 µm thick sections were cut from all TMA blocks using a microtome. Sections from the TMAs were either mounted onto ITO-coated conductive slides for MALDI MS analysis [57], or onto regular glass microscope slides for H&E staining. Paraffin removal was carried out using washes in xylene (100%, twice for 20 min) and graded ethanol washes (100% twice for 5 min and in successive washes in 95, 80 and 70% for 5 min each). The slides were then allowed to fully dry in an oven for 1 h at 65 °C. Antigen retrieval was performed by heating the section in a tris-EDTA buffer at pH 9 at 95°C for 20 min in a sealed vessel.

On-Tissue Digestion. Trypsin (100 µg) was dissolved in 200 µL of 50 mM acetic acid to obtain a stock solution with a concentration of 0.5 µg/µL. A 50 µL aliquot of this stock solution was activated by adding 500 µL of 100 mM ammonium bicarbonate to reach a pH of ~8 and a final trypsin concentration of 0.045 µg/µL. This trypsin solution was automatically spotted onto the TMA tissue sections

using a Portrait 630 reagent multi-spotter (Labcyte, Sunnyvale, CA) into an array incorporating 250 μm center to center spacing between individual spots, each of which were approximately 175 μm in diameter. The trypsin was spotted over a series of 30 iterations while depositing 1 drop (~160 pL per drop) each iteration to achieve a total spot volume of ~4.8 nL. Between each iteration, the deposited trypsin solution was allowed to dry (~5 min time intervals per iteration). The trypsin spotting proceeded at room temperature (~21°C) over a time period of ~2.5 h, allowing ample time for digestion to take place. Following digestion, a solution consisting of 10 mg/mL of CHCA in 1:1 ACN/0.5% TFA (aq.) was spotted directly onto the array of tryptic spots over 30 iterations at 1 drop per iteration.

Imaging Mass Spectrometry of Digested TMA Section. The printed arrays were analyzed using an Ultraflex II MALDI TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a smart beam laser [97] and controlled by the Flex Control 3.0 software package. The mass spectrometer was operated with positive polarity in reflectron mode and spectra acquired in the range of *m/z* 700-5000. Image acquisition of the spotted arrays was carried out using the Flex Imaging 2.0 (Bruker Daltonics, Billerica, MA) software package. A total of 1600 spectra were acquired at each spot position in a customized spiral raster pattern in 200 shot increments at a laser frequency of 200 Hz. The customized raster pattern was used to sample the entire spot area. Ion images were assembled using the Flex Imaging 2.0 software package.

MS/MS Sequence Analysis of Tryptic Peptides. MALDI MS/MS measurements of selected peptides were acquired using TOF/TOF fragmentation directly from the digested TMA sections. Each tryptic peptide sequenced was selected and fragmented manually and the generated spectra were processed in Flex Analysis 3.0 (Bruker Daltonics, Billerica, MA). Processing included 25 cycles of a Savitzky-Golay smoothing algorithm with a width of .15 and baseline subtraction using a median algorithm with a flatness value of .5 and a median value of .3. Monoisotopic peaks with a S/N > 8 were selected in each MS/MS spectrum using the SNAP peak picking algorithm. All MS/MS spectra were loaded into Biotools 3.0 (Bruker Daltonics, Billerica, MA) and converted into a single mascot generic format (.mgf) data file. This file was submitted into a MASCOT (Matrix Science, Boston MA) search engine and run against the Swiss-Prot database to match tryptic peptide sequences to their respective intact proteins. The MS/MS spectrum search was performed with a parent ion tolerance of 200 ppm and a fragment ion tolerance of ±0.4 Da. The search criteria also included up to 3 missed cleavages and variable modifications including, protein N-terminus acetylation, histidine/tryptophan oxidation, and methionine oxidation.

Data Analysis. Statistical analyses of MS profiles were carried out using ClinPro Tools 2.0 (Bruker Daltonics, Billerica, MA). Classes of spectra were loaded into the software and baseline correction was achieved using a top hat algorithm with a 10% minimal baseline width. ClinPro Tools automatically normalizes all spectra to their own total ion count (TIC). Thus, for each spectrum the TIC is

determined as the sum of intensities from all data points in the spectrum. Subsequently all data point intensities of this spectrum are divided by the obtained TIC value bringing all intensities into the range of [0,1]. Peaks in the spectra were selected manually and the maximum intensity within each of the defined peak integration areas was used as the comparative value. The classification model used in this analysis was built using a support vector machine algorithm (SVM) using 73 peaks determined through statistical comparison by means of a t-test and minimum three-fold intensity difference average. The number of neighbors for the K-NN classification parameter in the SVM settings was set to 3.

CHAPTER IV

HIGH-THROUGHPUT PROTEOMIC ANALYSIS AND STATISTICAL COMPARISON OF MULTIPLE FORMALIN-FIXED PARAFFIN-EMBEDDED LUNG CANCER TISSUE MICROARRAYS

Abstract

We have previously shown that *in situ* digestion coupled with imaging mass spectrometry can be used to map protein distributions in a large number of patient samples compiled into an FFPE-TMA. However, to elucidate the full potential of this technology for the proteomic analysis of FFPE clinical specimens it will need to be validated on a sufficiently large sample set. Here we use this method to analyze a set of four TMAs containing samples from 115 different NSCLC patients diagnosed with adenocarcinoma and squamous cell carcinoma. Statistical analysis of the peptide distributions is conducted to reveal peaks that can distinguish the adenocarcinoma patients from the squamous cell carcinoma patients with high sensitivity and specificity. Using these peaks, classification models are generated that can accurately identify these NSCLC sub-histologies. Several disease relevant proteins were identified directly from histopathological regions of interest in these analyses. Importantly, many of these proteins and distributions correlate with results previously reported in the literature. The data presented in this work reveals the potential of this method to aid pathologists in differentiating various tissue histologies and elucidating new cancer phenotypes.

Introduction

Pulmonary carcinomas are typically divided into two groups, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC is further divided into three main histological subtypes: squamous cell carcinoma (SCC), adenocarcinoma (AC), and large cell carcinoma (LCC). The accurate pathologic classification of lung cancers to distinguish NSCLC from SCLC is critically important as a prognostic factor and in the evaluation of treatment modalities.[120] For example, the biological characteristics and responsiveness to chemotherapy is profoundly different for patients with NSCLC (unresponsive) and SCLC (responsive). In many cases, pathologists are able to sufficiently distinguish these classes of lung cancer by evaluating the histological features of a biopsy with light microscopy. However, several factors including poorly differentiated cells, crush artifact, tumor necrosis, and limited tumor representation can preclude making a differential diagnosis using morphological examination alone.[120]

In the past, it was generally thought that specific sub-classification of NSCLC cases was unnecessary because all subtypes were considered to have the same prognosis and were therefore treated in a similar manner.[121] However, recent advances in targeted molecular based therapies have made the recognition and sub-typing of NSCLC increasingly important. For example, the anti-VEGF monoclonal antibody bevacizumab (Avastin; Genentech, South San Francisco, CA) is FDA approved for treatment of unresectable, locally advanced, recurrent, or metastatic, non-squamous NSCLC. However, patients with

squamous cell carcinoma cannot be administered Avastin because a 30% mortality rate has been observed in these individuals due to pulmonary hemorrhaging.[122, 123]

As with differentiating NSCLC and SCLC, making the distinction in NSCLC patients between glandular (AC) and squamous (SCC) differentiation can be challenging in poorly differentiated carcinomas, especially in small biopsy tissues obtained by bronchoscopy.[120] For example, Stang et al. examined the agreement in the histopathological evaluation of 688 lung cancer patient tissues between two independent pathologists. The observed agreement was 65% with a kappa value of .54 (95% CI: 0.49-0.58). The kappa statistic is an index which compares the agreement against that which might be expected by chance. When this value is broken down based on the different histologies examined, it is clear that the agreement is highest for the diagnosis of small cell carcinoma (0.94; $\kappa = 0.82$) and significantly lower for squamous cell carcinoma (0.81; $\kappa = 0.55$) and adenocarcinoma (0.81; $\kappa = 0.55$). Based on these values, the authors concluded that "if histological typing of lung cancer beyond the distinction between small-cell versus non-small-cell carcinoma is critical in a study, we would suggest to generally add ancillary techniques including histochemistry, immunohistochemistry, electron microscopy, and other molecular biological methods that may increase the reliability of the histopathological evaluation..." These limitations have motivated research initiatives which focus on identifying molecular markers that could aid pathologists in accurately making a differential diagnosis when morphological examination alone is not adequate.

In the previous chapter we demonstrated that in situ digestion followed by MALDI IMS analysis could effectively be used to map protein distributions in FFPE TMA samples. In a single experiment, this method was used to reproducibly detect hundreds of peptide peaks simultaneously in tissues from several different NSCLC patients. By combining the MS information with the histological assessment carried out by the pathologist, the resulting dataset was used to determine peak signatures that could sufficiently distinguish the adenocarcinoma and squamous cell carcinoma patients. The potential of this technology was clearly established, however further validation of its applicability to clinical samples will require analysis of a large number of patient samples across multiple TMAs. Here we apply this method for the proteomic analysis of a set of cross institutional TMAs containing a total of 115 different NSCLC patients. Peak distributions unique to the different cancer subtypes are identified through statistical comparison and used to build classification models. The biological relevance of the proteins detected in these samples is discussed and compared to literature reports.

Results

In situ tryptic digestion followed by MALDI IMS was conducted on four separate FFPE lung cancer TMAs. A total of 160 separate tissues were analyzed in these experiments from 115 NSCLC patients, including 52 diagnosed with adenocarcinoma and 63 diagnosed with squamous cell carcinoma. For each TMA section analyzed by mass spectrometry, a serial section was collected

and stained with H&E. The stained sections were reviewed by a pathologist and the relevant pathological regions of each tissue were marked accordingly. These annotated images were then overlaid with the sections analyzed using mass spectrometry enabling each mass spectrum to be assigned to a specific patient and histological region.

Spectra acquired from regions marked as either adenocarcinoma or squamous cell carcinomas were grouped together for statistical analysis. Each spectrum contains several hundred peaks representing a diverse population of tryptic peptides. The monoisotopic peaks present in each spectrum were identified and extracted as a peak list containing m/z values and their respective intensities. For statistical analysis, all spectra acquired from each individual patient were averaged into a single spectrum and peak distributions in these average spectra were then compared.

A total of 335 unique monoisotopic peaks were detected in these datasets using a S/N 5 cutoff. It is noted that this value increases to 602 when a S/N of 3 cutoff is used. For statistical analysis and classification, it was determined that the set of 335 peaks included a sufficient number of classifiers to distinguish the two cancer classes and that use of the larger peak lists (602) did not increase the accuracy. However, it is likely that the peak information contained in the expanded peak lists generated using a lower S/N threshold does contain valuable information and may be especially useful in revealing subtypes within a specific histology.

Significance analysis of microarrays (SAM) was conducted on the set of 335 peaks to determine m/z values that were significantly different between the two cancers and therefore could be used to distinguish the patient classes. Briefly, SAM is a statistical technique originally developed for determining significant genes in a set of microarray experiments.[124] This method can be easily adapted to evaluate protein/peptide expression data in a set of patients because the general concept of the two analyses is the same, in that an expression level of a specific attribute (gene or protein) is measured. The input to SAM is the peak intensity values across all patients, as well as a class defining variable, which in these studies was either squamous cell carcinoma or adenocarcinoma. SAM conducted a two-class unpaired comparison using a t-statistic over 5000 permutations. SAM permutes the data by scrambling the label of samples and therefore assumes that all null hypotheses are true and there are no differences in the peptide expression levels between the two cancer classes. For each permutation, SAM assigns a score of relative difference for each peak and ranks these by their magnitude. This value is compared to the expected relative difference under the null hypothesis, which is defined as the average of the ranked scores of relative difference over all permuted data. The peaks that show relative difference levels deviating from the expected value under the null hypothesis by an amount greater than a thresholding parameter, Δ, are considered significant. A minimum fold intensity change value is also used to define the significance cutoff value.

The SAM comparison determined a set of 84 peaks out of the set of 335 that were significantly different between the adenocarcinoma and squamous cell carcinoma patients with a minimum 2-fold intensity change and a false discovery rate <1 **(Figure 27)**. In terms of distribution, 58 of these markers were significantly higher in the squamous cell carcinoma patients and 26 were higher in the adenocarcinoma patients.

From this set of 84 peaks, a subset of 29 peaks was determined to be optimal for distinguishing the two histological classes using a feature selection function and was used to build an SVM classification model. However, through careful evaluation of the peaks present across the patients in these datasets it became clear that a small percentage of patients diagnosed with adenocarcinoma exhibited peak distributions highly consistent with squamous cell carcinoma patients and vice versa. In other words, the spectra from these patients' showed high intensity values for the peaks determined through SAM to be significant for the cancer subtype contrary to their diagnosis. This is clearly seen when all patients are clustered based on these 29 peaks, as shown in **Figure 28**. An even smaller percentage of patients did not express (above threshold) any of the 29 peaks used in the model. In total, 10 patients (3 adenocarcinoma and 7 squamous cell carcinoma) exhibited peak distributions consistent with the opposite class and 6 patients were unclassifiable using these
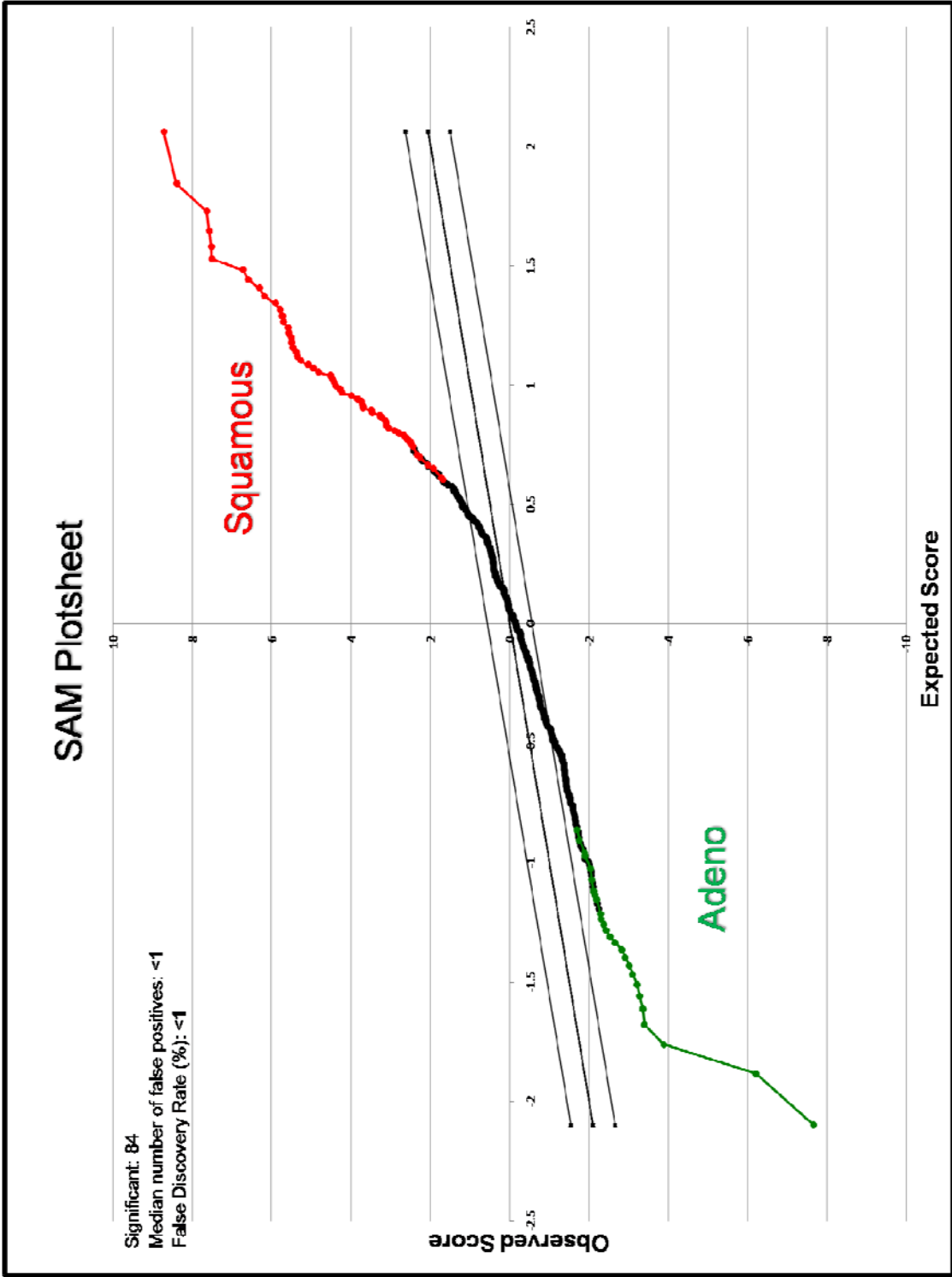
Figure 27. SAM plot showing the distribution of all peaks including those determined to be significant for a specific histology.

91
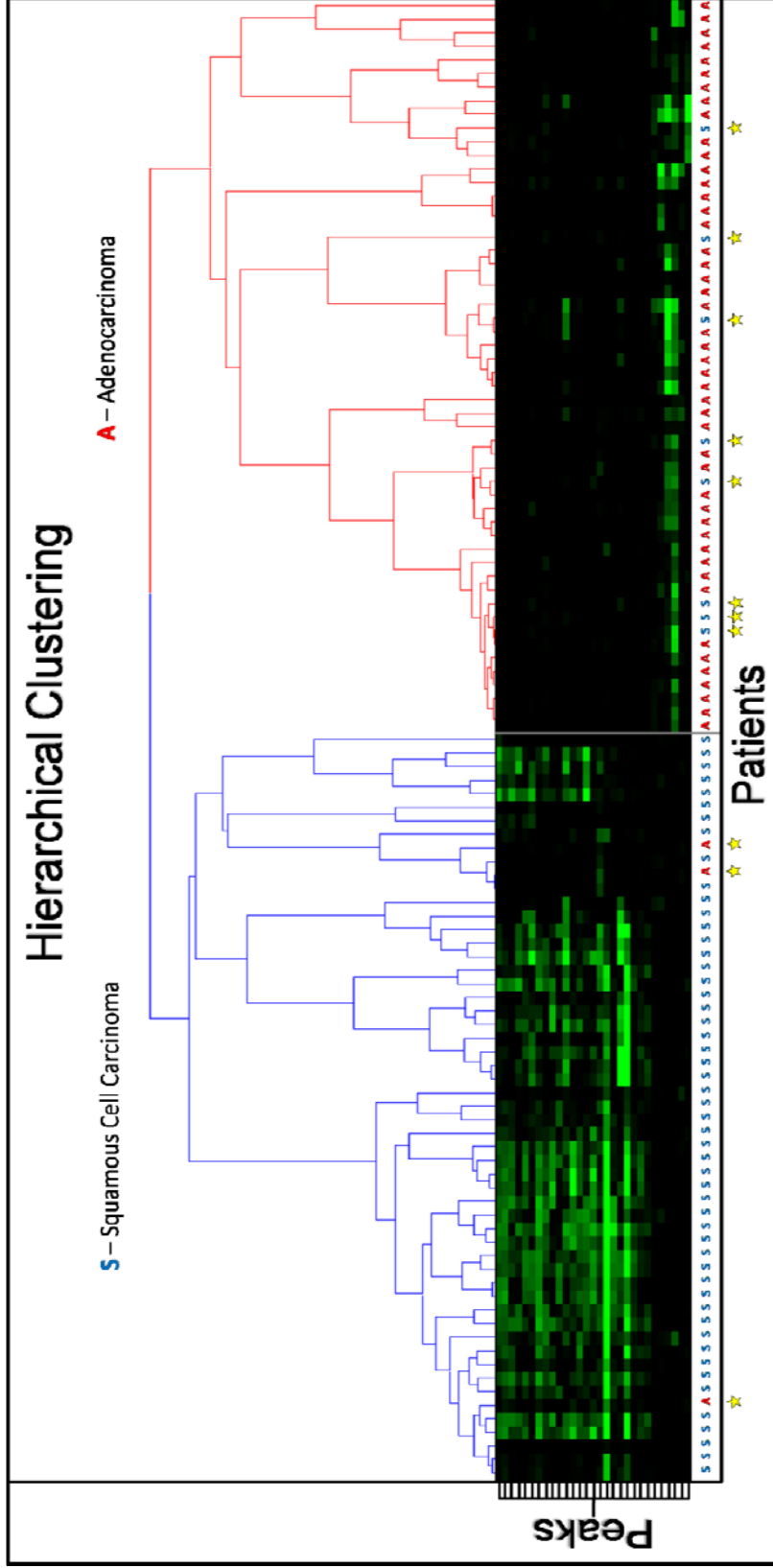
Figure 28. Hierarchical clustering of all patients using the 29 peaks determined to be optimal for classification of the squamous cell carcinoma and adenocarcinoma patients. The stars indicated patients that showed peak distributions that contradicted the diagnosis made by the pathologist.

peaks.   Therefore, these outlier patients were excluded from the model generation and classification steps.

Using these 29 peaks, a set of SVM classification models were built that could classify the remaining 99 patients with a high degree of accuracy, as shown in **Table 2**. Models **a)** and **b)** are based on an iterative train/test workflow where the dataset is split into a training set and a test set. The SVM model is built on a training set using the 29 peaks then used to classify the remaining test set. For model **a)** the training set consisted of 80% of the patients and the remaining 20% were used as the test set. This splitting process for the training and test sets was randomly repeated 100 times and the average classification accuracy using this model was 99.3%. To further demonstrate the robustness of these 29 peaks as classifiers, a second SVM model, **b)**, was built using a 50/50 split for the training and test sets and the average classification accuracy over 100 repetitions was 98.2%.   Model **c)** was built using a slightly different approach, where the SVM was trained using a set of adenocarcinoma and squamous cell carcinoma template spectra.   The template spectra were built based on the average value for each peak across all patients from that class. The SVM model was then generated using these template spectra and used to classify the entire dataset. This model was able to classify the patients with 100% accuracy.

The template model was also used to conduct a spectrum based classification, where each individual spectrum from each patient was classified without averaging. The results of this analysis are outlined in **d).**

Table 2. Results from the SVM classification using four separate classification workflows.

a.) Support Vector Machine Classifier (80/20 Train/Test)

Train-test parameters		Confusion matrix				Accuracy	
			Adeno	Squamous	Sum		% Correct
Train proportion (%)	75	Adeno	927	0	927	Adeno	100
Trials	100	Squamous	15	1058	1073	Squamous	98.6
Peaks Used	29	Sum	942	1058	2000	Overall	99.3
Patients	99						

b.) Support Vector Machine Classifier (50/50 Train/Test)

Train-test parameters		Confusion matrix				Accuracy	
			Adeno	Squamous	Sum		% Correct
Train proportion (%)	50	Adeno	2306	41	2347	Adeno	98.3
Trials	100	Squamous	48	2605	2653	Squamous	98.1
Peaks Used	29	Sum	2354	2646	5000	Overall	98.2
Patients	99						

c.) Support Vector Machine Classifier (Template Based)

Train-test parameters		Confusion matrix				Accuracy	
			Adeno	Squamous	Sum		% Correct
Template spectra used to train model	1	Patients					
Trials	1	Adeno	46	0	46	Adeno	100
Peaks Used	29	Squamous	0	53	53	Squamous	100
Patients	99	Sum	46	53	99	Overall	100

d.) Support Vector Machine Classifier (Template Based - Individual Spectra Classification)

	Patients	Total Spectra	Correct	Incorrect	%
Adenocarcinoma	46	1362	1347	15	98.9
Squamous Cell Carcinoma	53	1885	1853	32	98.3
Totals	99	3247	3200	47	98.6

Discussion

Here we show that the analysis of FFPE TMAs using on-tissue digestion coupled with MALDI IMS can simultaneously reveal the distribution of numerous proteins across a large number of patient samples. One important aspect of this technology is that it can be applied to tissues acquired at different institutions, and therefore assembled into different TMA blocks, and the results generated from these separate analyses can be effectively compared. We applied this methodology to four separate TMAs including two containing NSCLC patient samples from Vanderbilt University Medical Center and two assembled by an independent biotechnology company (Folio Biosciences; Columbus, OH).

Several hundred different peptide distributions were detected in these samples using IMS. By combining the MS information with the histological evaluation carried out by the pathologist, these distributions can be linked to specific cellular regions of interest. This workflow can discover peak patterns that are specific to the squamous cell carcinoma or adenocarcinoma subtypes. These peak distributions are then used to develop statistical classification models capable of assigning an unknown spectrum from a test set to the appropriate class.

As outlined in **Table 2,** the models generated were able to classify the selected test data with remarkable accuracy using 3 different approaches. However, the reported accuracies were contingent upon the removal of a relatively small subset of patient outliers from each class. These patients were removed from the dataset prior to conducting the model train/test workflow

because they either exhibited peak profiles consistent with the histology contrary to that reported during the histological evaluation by the pathologist or they did not express any of the peaks used for classification above the threshold levels.

A number of factors may explain this observation. For one, the diagnoses made by the pathologist, to which the observed peak distributions are compared to, are not necessarily a "gold standard". As mentioned in the introduction, several studies comparing the diagnoses made by independent pathologists on NSCLC tissues reported a high level of disagreement. The subjectivity of using morphological examination alone is likely compounded by factors such as poorly differentiated cells and limited tissue quantities. Our method provides an objective measurement of protein expression in these patients and therefore should not be limited by factors such as a small quantity of tissue or undifferentiated cells (provided the underlying biology has not changed). This assertion is backed by the fact that the vast majority of patients in each class exhibit a peak profile that can be accurately identified using a panel of 29 peaks. However, it is possible that some tumors exhibiting cellular differentiation attributed to a specific NSCLC histological subtype, such as adenocarcinoma or squamous cell carcinoma, may exhibit proteins that are typically unique to a different histology. This appears to be the case for the patient outliers in our dataset seeing that the histological diagnosis assigned to these tissues were agreed upon by at least two lung cancer experts.

These results are very interesting, though not completely unexpected when considering the well-known heterogeneity of lung cancer.[125] The

hierarchical clustering displayed in **Figure 28** clearly shows the presence of subgroups within the major histological subtypes of squamous cell carcinoma and adenocarcinoma. This elucidates the complexity involved in the classification of squamous cell carcinoma and adenocarcinoma subhistologies in NSCLC and further highlights the potential of using MS data to go beyond a simple two class distinction. Determining the relevance of these subgroups in terms of clinical characteristics will require analysis of a large patient cohort where detailed treatment and follow-up information is available.

The amount of information contained in these NSCLC patient datasets is extraordinary. The combined MS and histological analysis provides an optimal platform to compare proteomic distributions between the patients. For example, many of the proteins identified in these analyses are from the cytokeratin family. Cytokeratins (CK) are intermediate filament proteins which make up part of the integrated cytoskeleton of both normal and malignant epithelial cells.[126] Cytokeratins are grouped into a type I (acidic, CK9 through CK20) and a type II (neutral basic, CK1-CK8) gene family. The individual cytokeratin proteins assemble into non-covalent heterodimers containing one type I species and one type II species. These heterodimers are then organized further into filamentous structures that in turn provide structural and mechanical stability to cells and their respective tissues.[126] The expression of cytokeratins in cells is remarkably tissue specific and differentiation dependent, suggesting that the pattern of cytokeratin species present in cells is related to their biological function.[127, 128]

Cytokeratin 8 (CK8) and 18 (CK18) are the major components of all simple or single layer epithelial tissues. The carcinomas that arise from these tissues typically continue to express CK8 and CK18 and therefore these proteins are commonly used as markers of epithelial derived cancers. Given the specialized functions of CK8 and CK18 in the tissues of different organs, such as albumin production in liver cells and casein production in mature mammary gland cells, it would be expected that their expression would discontinue in parallel to other differentiated functions during malignant conversion and progression.[129] However, detection of CK8 and CK18 has commonly been reported in both well and poorly differentiated carcinomas. Oshima et al.[129] thoroughly review the oncogenic regulation and function of CK8 and CK18 and provide many interesting mechanisms behind this continued expression. In addition to these considerations, several reports have shown the presence of CK8/CK18 in tumors derived from tissues that do not naturally express these simple epithelial keratins. For example, CK8 and CK18 are not found in stratified epithelia. However, squamous cell carcinomas originating from stratified epithelia do show anomalous expression of the CK 8/18 pair.[129-132]

In accordance with the aforementioned IHC studies, CK 8 and CK 18 are co-expressed in many of the lung tumor biopsies analyzed in these studies, including several cases of squamous cell carcinoma. For example, **Figure 29** shows the corresponding distributions of a CK 8 peptide and a CK 18 peptide in 3 squamous cell carcinoma biopsies and 1 adenocarcinoma biopsy from a lung
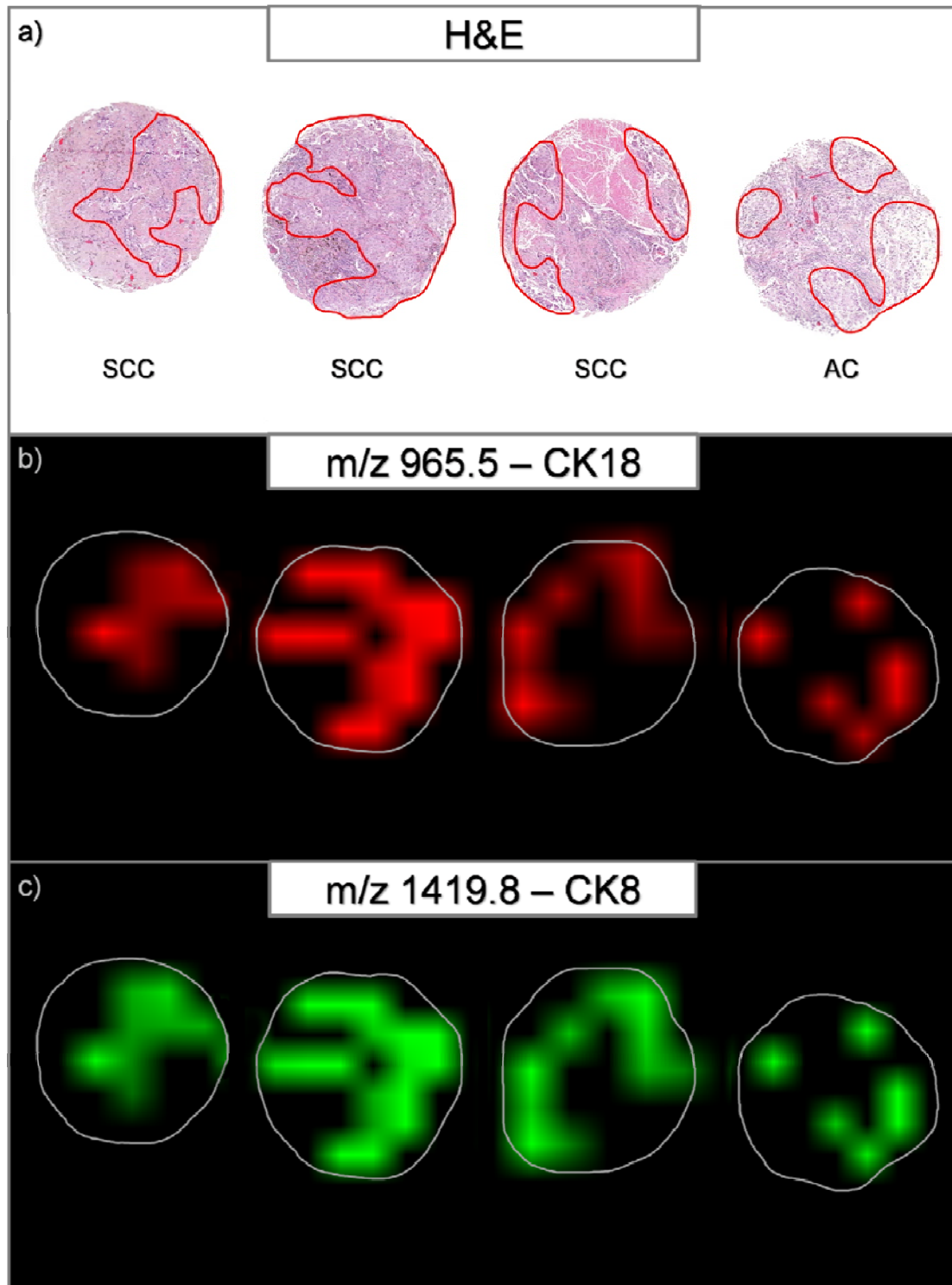
Figure 29 a) H&E stained tissues from 3 squamous cell carcinoma (SCC) patients and an adenocarcinoma (AC) patient. b) and c) show the co-expression of a peptide from CK18 and a peptide from CK8 across the different patient samples.

cancer FFPE-TMA. Currently, we have identified five peptides specific for CK 8 and four peptides specific for CK 18 in these tissues. Interestingly, in correlation with the IHC studies, three different patterns of CK 8 and CK 18 expression are present in this dataset: **1.)** CK 8+/CK 18+ **2.)** CK 8-/CK 18+ **3.)** CK 8+/CK 18-. Recently, these proteins have been associated with several important tumor characteristics including invasiveness, metastasis, and drug resistance.[133] The ability of this method to quickly and specifically detect their presence and map their distributions in tissues will facilitate future studies of these proteins.

In addition to CK8 and CK18, we have identified 12 different cytokeratin species in the lung cancer TMAs. The results of the SAM evaluation clearly illustrate the presence of histology dependent cytokeratin distributions with distinctly different levels of expression between the two cancer classes. For example, peptides from cytokeratin 5 and 6 (CK5 and CK6) showed highly increased expression levels in the squamous cell carcinoma patients. In terms of patients, detectable levels of CK5 were present in 60 (88%) squamous cell carcinoma patients and 27 (51%) adenocarcinoma patients and CK6 in 43 (63%) squamous cell carcinoma patients and 2 (3%) adenocarcinoma patients.

These results are interesting when considering the literature reports on the distributions of CK5 and CK6. CK 5/6 is widely used in modern surgical pathology because it characteristically stains squamous carcinomas strongly and diffusively. A recent literature report[134] argued that a lung tumor that is not CK 5/6 positive is not squamous cell carcinoma, although this study was carried out on a relatively small set of patients (45 NSCLCs: 24 adenocarcinomas and 21

100

squamous cell carcinomas). Numerous other studies have also shown the ability to use a CK 5/6 antibody for distinguishing pulmonary squamous cell carcinomas from adenocarcinomas, yet the reported specificity of this marker in distinguishing these histologies varies significantly between studies.[120, 134-138]

We conducted a comparison between the peptide distributions from these proteins and the IHC distributions using the CK 5/6 antibody. One important aspect of the antibody used in the abovementioned studies and our analyses is that it is a mixture of mouse monoclonal antibodies raised against purified CK proteins that shows non-specific reactivity against CK5 and CK6. For clarification in the following figure, it should be noted that the TMA used in these experiments consists of triplicate needle core biopsies from each patient and therefore the layout is such that each row contains 9 tissue cores (3 cores from 3 different patients). Interestingly, when the localizations of CK5 and CK6 are mapped using peptides unique to each protein it is clear that their distributions are not identical. For example, **Figure 30 a.)** shows the distribution of the CK 5/6 antibody using IHC on triplicate biopsies from two patients diagnosed with squamous cell carcinoma. Figures c and b show the distributions of a peptide specific for CK5 and CK6, respectively. These ion density maps show that the distributions of CK5 (m/z 1410.7) and CK6 (m/z 1407.7) are not identical, yet due to the non-specific binding of the CK5/6 antibody this cannot be readily determined using IHC.
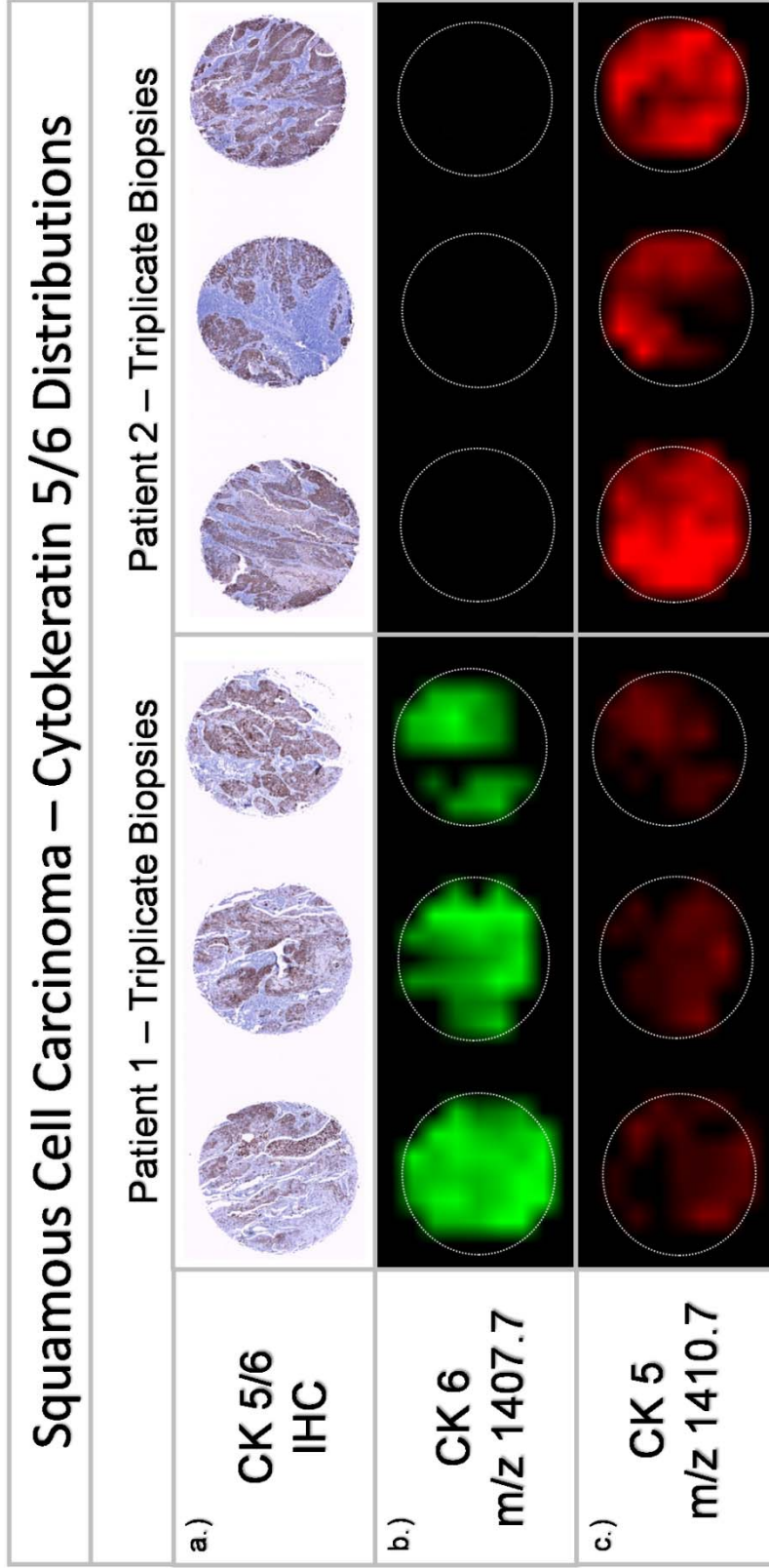
Figure 30. a) IHC using CK 5/6 antibody on triplicate squamous cell carcinoma cores from two patients. **b)** Distribution of a peptide identified to be specific for CK6. **c)** Distribution of a peptide identified to be specific for CK5.

These results are significant for several reasons. First, it is important to show that the general ion distributions observed in IMS experiments can be validated using alternative methods such as IHC. Furthermore, this scenario highlights the superior molecular specificity that can be achieved using mass spectrometry as compared to other imaging methods.

Other ions, such as peptides from the protein cytokeratin 7 (CK7) showed intensity fold-change values of up to 15 times higher in the adenocarcinoma patients. Detectable levels of CK7 were present in 49 (95%) of the adenocarcinoma patients and 22 (33%) of the squamous cell carcinoma patients.

CK7 is commonly used in pathology in combination with CK20 to discriminate between metastatic and primary tumors from several different organs.[139] This highlights another potential application of this method: using the protein expression patterns detected in a metastatic tumor to determine primary sites of origin. Unfortunately, metastatic cancer with an unknown primary is one of the 10 most common cancer diagnoses, with 3-5% of all cancer patients presenting with this condition.[140] Determining the primary site of tumor origin is important for disease management and prognosis in that a more precise cancer diagnosis will lead to a more effective course of treatment and ultimately improve the overall outcome.

Adenocarcinomas are the most common malignant neoplasms of unknown primary origin, making up about 60% of all cases.[140] Determining the tissue origin of metastatic adenocarcinoma can be extremely difficult because the microscopic appearance of histological features in these neoplasms is not

distinctive enough to make a confident diagnosis. Therefore, a panel of molecular markers is needed to assist in evaluating these tumors and identifying their tissue of origin. This method has the ability to quickly detect a phenotype in a tissue and therefore could potentially be an extremely powerful way to discern the primary sites for malignant tumors.

Cytokeratin fragments detected in the sera of cancer patients have shown promise to be a simple, cheap, non-invasive and reliable tool for determining prognosis and effectiveness of therapy for certain cancers. The mechanism responsible for the release of cytokeratins or cytokeratin fragments into the serum is not completely understood, but is thought to involve multiple pathways including proteolytic degradation from dying cells, abnormal mitosis, spillover of monomeric cytokeratin peptides from proliferating cells, apoptosis, etc.[126, 141-143] Cytokeratin fragments have been detected in a number of body fluids including blood, urine, cyst fluids, ascites, pleural effusions, and cerebrospinal fluid following release from cancer cells.[126] The fact that in normal, apparently healthy individuals, the level of cytokeratin fragments detected in circulation is low and increases significantly in patients with carcinomas highlights the significance of these results.

Currently, there are three cytokeratin markers which are commonly used as prognostic markers in lung cancer: tissue polypeptide antigen (TPA) for CK8, 18, and19; tissue polypeptide specific antigen (TPS) for CK18; cytokeratin fragments 21-1 (Cyfra 21-1) and CK19. The level of these markers in serum is quantified using various commercially available serological assays. Analysis of

the detected levels of these markers in serum has been demonstrated to assist in the early detection of recurrence and to enable a rapid assessment of a patient's response to therapy.

Several studies have reported that these cytokeratins are prognostically significant markers in patients with NSCLC.[144-146] Most notably, Cyfra 21-1has been shown to be an independent predictor of prognosis in NSCLC [147-150], the most widely used serum marker for screening and monitoring of lung cancer[145, 149, 151-155], post therapeutic evaluation [153, 156], and indicator of advanced disease.[155, 157, 158]

Although promising, the current set of serum tumor markers lack the reliable specificity and sensitivity necessary to gain wide acceptance as clinical tools. The clinically relevant data that has been collected using a relatively small set of tumor markers makes it reasonable to assume that there are additional tumor markers present that have yet to be discovered. As discussed previously (Chapter 1), the extreme complexity of blood makes discovery of tumor markers difficult. This limitation has encouraged research focusing on the direct analysis of the resected tumor or tumor biopsy to carry out the biomarker discovery phase. It is logical to hypothesize that the concentration of potential biomarkers will be highest in the tumor and its adjacent microenvironment.

Our work has revealed a complex pattern of cytokeratin expression in NSCLC tumors that may account for the inconsistencies in the literature reports regarding these tumor markers. It would be valuable to assess the correlation between the levels of these markers in a patient's serum and the tumor tissue.

Most of the previously reported studies looking at NSCLC serum biomarkers have been driven under the assumption that these tumors should express similar markers during progression. However, our results indicate that the protein expression patterns of these tumors can be markedly different and it is therefore likely that a panel of markers will need to be employed to sufficiently detect these tumors.

In addition to cytokeratins, many other proteins were identified that showed histology specific distributions. Several of these proteins have previously been reported to play a role in lung cancer including apolipoprotein A-I[159], carbonyl reductase[160], carcinoembryonic antigen[161], aldo-keto reductase family 1 C1[162], alpha enolase[162], glutathione s-transferase[162], and annexin A2[163] (**Figure 31**). These images were all generated simultaneously in a single imaging experiment. This figure helps put into perspective the extraordinary amount of information that can be generated in a single FFPE-TMA imaging experiment when it is considered that to generate similar information using IHC, seven separate experiments would have to be conducted including controls.
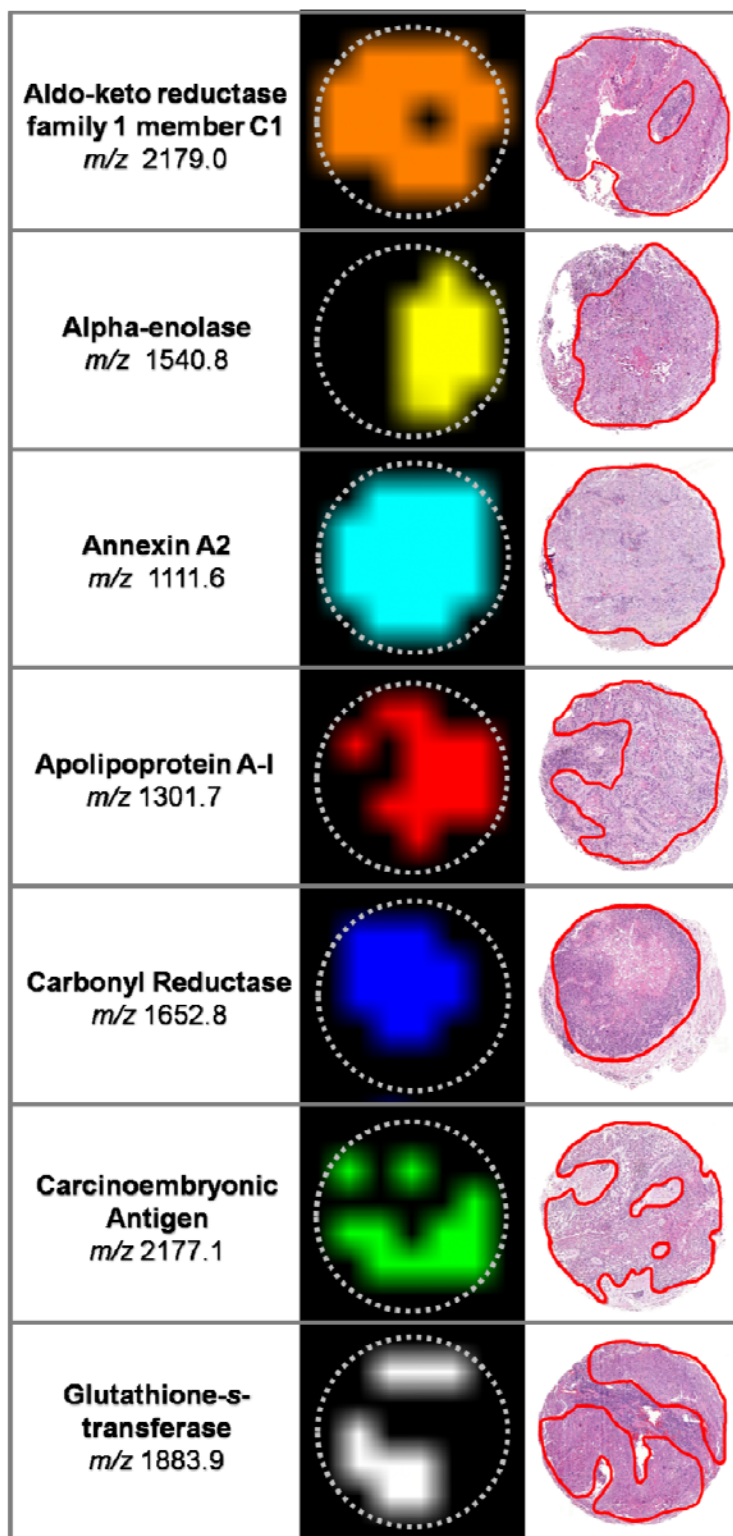
Figure 31. Distribution of the peptides identified from 7 different proteins that have previously been associated with NSCLC.

Materials and Methods

Materials. HPLC grade acetonitrile (ACN) and trifluoroacetic acid (TFA) were purchased from Fisher Scientific (Pittsburgh, PA). Alpha-cyano-4-hydroxy-cinnamic acid (CHCA) was purchased from Fluka (Buchs, Switzerland) and used without further purification. Trypsin Gold was purchased from Promega (Madison, WI). ITO-coated conductive slides were purchased from Delta-Technologies (Stillwater, MN). Four TMAs containing needle core biopsies from NSCLC patients were used in these studies. Two of the TMAs were assembled at Vanderbilt University Medical Center and two were assembled by an independent biotech company (Folio Biosciences, Columbus OH). These four TMAs consisted of a total of 157 tissues from 115 different patients diagnosed with squamous cell carcinoma or adenocarcinoma by a pathologist. The tissue cores in the Folio Biosciences TMAs were 1.5 mm in diameter and the cores in the Vanderbilt TMAs were 1 mm in diameter.

Tissue Preparation. Serial 5 µm thick sections were cut from all TMA blocks using a microtome. Sections from the TMAs were either mounted onto ITO-coated conductive slides for MALDI MS analysis [57], or onto regular glass microscope slides for H&E staining. Paraffin removal was carried out using washes in xylene (100%, twice for 20 min) and graded ethanol washes (100% twice for 5 min and in successive washes in 95, 80 and 70% for 5 min each). The slides were then allowed to fully dry in an oven for 1 h at 65 °C. Antigen retrieval

was performed by heating the section in a tris-EDTA buffer at pH 9 at 95°C for 20 min.

On-Tissue Digestion. Trypsin (100 µg) was dissolved in 200 µL of 50 mM acetic acid to obtain a stock solution with a concentration of 0.5 µg/µL. A 50 µL aliquot of this stock solution was activated by adding 500 µL of 100 mM ammonium bicarbonate to reach a pH of ~8 and a final trypsin concentration of 0.045 µg/µL. This trypsin solution was automatically spotted onto the TMA tissue sections using a Portrait 630 reagent multi-spotter (Labcyte, Sunnyvale, CA) into an array incorporating 250 µm center to center spacing between individual spots, each of which were approximately 175 µm in diameter. The trypsin was spotted over a series of 30 iterations while depositing 1 drop (~160 pL per drop) each iteration to achieve a total spot volume of ~4.8 nL. Between each iteration, the deposited trypsin solution was allowed to dry (~5 min time intervals per iteration). The trypsin spotting proceeded at room temperature (~21°C) over a time period of ~2.5 h, allowing ample time for digestion to take place. Following digestion, a solution consisting of 10 mg/mL of CHCA in 1:1 ACN/0.5% TFA (aq.) was spotted directly onto the array of tryptic spots over 30 iterations at 1 drop per iteration.

Imaging Mass Spectrometry of Digested TMA Section. The printed arrays were analyzed using an Ultraflex II MALDI TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a smart beam laser [97] and controlled by

the Flex Control 3.0 software package. The mass spectrometer was operated with positive polarity in reflectron mode and spectra acquired in the range of *m/z* 700-5000. Image acquisition of the spotted arrays was carried out using the Flex Imaging 2.0 (Bruker Daltonics, Billerica, MA) software package. A total of 1600 spectra were acquired at each spot position in a customized spiral raster pattern in 200 shot increments at a laser frequency of 200 Hz. The customized raster pattern was used to sample the entire spot area. Ion images were assembled using the Flex Imaging 2.0 software package.

MS/MS Sequence Analysis of Tryptic Peptides. MALDI MS/MS measurements of selected peptides were acquired using TOF/TOF fragmentation directly from the digested TMA sections. Each tryptic peptide sequenced was selected and fragmented manually and the generated spectra were processed in Flex Analysis 3.0 (Bruker Daltonics, Billerica, MA). Processing included 25 cycles of a Savitzky-Golay smoothing algorithm with a width of .15 and baseline subtraction using a median algorithm with a flatness value of .5 and a median value of .3. Monoisotopic peaks with a S/N > 8 were selected in each MS/MS spectrum using the SNAP peak picking algorithm. All MS/MS spectra were loaded into Biotools 3.0 (Bruker Daltonics, Billerica, MA) and converted into a single mascot generic format (.mgf) data file. This file was submitted into a MASCOT (Matrix Science, Boston MA) search engine and run against the Swiss-Prot database to match tryptic peptide sequences to their respective intact proteins. The MS/MS spectrum search was performed with a parent ion tolerance

of 200 ppm and a fragment ion tolerance of ±0.4 Da. The search criteria also included up to 3 missed cleavages and variable modifications including, protein N-terminus acetylation, histidine/tryptophan oxidation, and methionine oxidation.

Spectra Processing. All spectra were processed using FlexAnalysis 3.0 (Bruker Daltonics; Billerica, MA). Baseline subtraction was carried out using the median algorithm with a flatness of .9 and a median level of .5. The Savitzky-Golay smoothing algorithm was used with a width of .2 m/z and 15 cycles. Monoisotopic peak picking was performed using the SNAP algorithm with a S/N cutoff of 5 and a quality factor threshold of 150. Within the processing script a function was included to perform a statistical internal calibration (mass error Δ 50 ppm) and a background peak removal based on a custom list of pre-determined peaks (e.g. trypsin autolysis peaks).

Data Processing. The monoisotopic peak lists were exported from the processed spectra using in-house software (Peak List Exporter, Vanderbilt MSRC Bioinformatics Core). These peak lists were then zero-padded (m/z 700-3000) at a bin interval of .1 Da to create full spectra containing only the monoisotopic peaks chosen by the processing software. Arrayed peak list spreadsheets were generated for all patients using ClinPro Tools 2.0 (Bruker Daltonics, Billerica, MA). All spectra from each individual patient were loaded into the software as a single class and the peak statistic export function was used to generate a spreadsheet containing all peaks and the respective average

intensity values of these peaks for each of the 115 patients. ClinPro Tools automatically normalizes all spectra to their own total ion count (TIC). Thus, for each spectrum the TIC is determined as the sum of intensities from all data points in the spectrum. Subsequently all data point intensities of this spectrum are divided by the obtained TIC value bringing all intensities into the range of [0,1].

Statistical Analysis. Tanagra data mining software[164] was used to build and evaluate the statistical classification models. Support vector machine models were built using the following parameter settings: Kernel: *Normalized Polynomial*; Polynomial exponent: *3*; Gamma for RBF: *0.01*; Complexity parameter: *1.00*; Attribute transformation: *Normalization*; Kernel for rounding: *1 x 10^{-12}*; Tolerance for accuracy: *.001*. Using the instance selection feature two different types of models were built using the train/test function to conduct an iterative training and testing workflow. The models were generated using different proportions to train the SVM model (80% and 50%) and the remainder of the dataset was used to evaluate the accuracy. A third model was built using template spectra for the adenocarcinoma and squamous cell carcinoma phenotypes. These spectra consisted of the average intensity values for each peak across the respective class of patients. This model was then used to classify all patients in the dataset. Hierarchical clustering was carried out using the software program Hierarchical Clustering Explorer 3.5 (HCE 3.5).[165] The data was clustered without

normalization using the complete linkage method and the Pearson correlation coefficient similarity/distance measure.

CK 5/6 Immunohistochemistry. Five micron TMA sections were placed on charged slides. The sections were rehydrated and placed in a heated Target Retrieval Solution (Labvision, Fremont, CA). Endogenous peroxidase was neutralized with 0.03% hydrogen peroxide followed by a casein-based protein block (DakoCytomation, Carpinteria, CA) to minimize nonspecific staining. The sections were incubated with Cytokeratin 5/6 (Catalog ab17133, abcam, Cambridge, MA) diluted 1:50 overnight. The Dako Envision+ HRP/DAB System (DakoCytomation) was used to produce localized, visible staining. The slides were lightly counterstained with Mayer's hematoxylin, dehydrated and coverslipped.

Chapter V

CONCLUSIONS AND PERSPECTIVES

A single thin tissue section contains an enormous amount of molecular information including proteins, peptides, lipids, and small molecules. These basic entities of biological systems are further arranged into an extensive and complex network of spatial arrangements governed by the function and physiological state of the underlying cellular components. IMS has the extraordinary ability to effectively detect a wide range of these biomolecules from a tissue with minimal sample preparation while maintaining the spatial orientation. However, in its current state, this technology is only capable of detecting a subset of species from these classes of molecules. There are several reasons for these limitations ranging from the physical and chemical properties of different molecular species (e.g. ionization efficiency, solubility, molecular weight...etc.), to the design and detection capability of the mass spectrometers used in these analyses. Therefore, to expand the applicability of this technology to a broader assortment of molecules, novel methodologies need to be developed.

One way to enhance the range of detectable species is to implement chemical reactions on the tissue prior to the standard sample preparation protocols. Chapter 2 outlines the development of a technique called *in situ* digestion, where an enzyme is used to digest the proteins at precisely defined

locations in a tissue prior to MS analysis. We showed that this process can be carried out efficiently and reproducibly using robotic spotters. The advantages and contributions of *in situ* digestion to this technology are several fold: 1) it enables high molecular weight proteins > 30 kDa to be indirectly mapped by detecting the distribution of their tryptic peptides; 2) proteins can be identified directly from a tissue by MS/MS sequence analysis of the tryptic peptides; 3) It enables the analysis of FFPE tissues opening the door to a vast archive of clinical samples.

In summary, the work outlined in Chapter 2 shows the feasibility of carrying out on-tissue chemistry prior to an MS analysis to expand the proteomic information detected. Our experiments focused on the use of trypsin; however a number of other proteases were successfully tested and could potentially be used to target different subsets of analytes. Furthermore, we were able to use a combination of enzymes, PNGase F and trypsin, to cleave and identify the carbohydrate constituent of the glycoprotein Copeptin. These results highlight the enormous potential that exists to further develop novel tissue chemical treatment methodologies prior to IMS analysis.

Experiments involving FFPE tissues are complicated by the very nature of the fixation process itself, in that the chemical modifications that preserve these samples render them unsuitable for routine proteomic analysis. The development of antigen retrieval techniques over the past decade was a vital step in making these samples accessible to different molecular analysis techniques including mass spectrometry. However, in terms of IMS experiments

where a tissue section is kept intact, it is highly unlikely that antigen retrieval completely de-crosslinks the proteins. This makes protein extraction directly from the surface of a tissue difficult and helps explain the fact that protein imaging experiments conducted on antigen retrieved FFPE sections have generally produced low quality data. Interestingly, we show in Chapter 3 and 4 that *in situ* digestion coupled with IMS has the ability to generate high quality tryptic peptide profiles comparable to those observed using fresh frozen tissues. Without antigen retrieval, in situ digestion is not possible, in that no tryptic peptides are observed in the MS analysis. This seems to indicate that the antigen retrieval step serves to provide a partial protein de-crosslinking as well as a denaturing effect that enables the enzyme access to carry out digestion. Further systematic studies are needed to elucidate the details of this mechanism but our results show that these analyses can be carried out reproducibly and provide meaningful data.

One distinct advantage of working with FFPE specimens is the ability to compile numerous patient tissues into a single TMA. A TMA takes a relatively small representative region from each individual patient and enables a high-throughput comparison to be performed. This provides an unparalleled platform for examining different protein distributions between patients in a large sample cohort. Working with a TMA, as opposed to each patient sample individually, avoids the effects of day-to-day fluctuations in the sample preparation and mass spectrometric analysis steps. It can be assumed that each sample in the TMA underwent nearly identical preparation (e.g. trypsin/ matrix application) and that

the working order of the mass spectrometer was constant during the analysis. Analysis of a large number of individual patient samples may take several days or even weeks to complete. This is an important consideration due to the sensitivity of the mass spectrometer, where even slight fluctuations in sample preparation or instrument status can create bias in the acquired data.

However, there are cases, including analysis of highly heterogeneous samples, where it may be necessary to acquire data from the entire patient tissue to accurately assess the molecular profile. In general though, the typical 500-1500 μm core taken for a TMA should provide a sufficient representation of the bulk sample, and usually duplicate cores are taken from each sample in order to evaluate the consistency.

In summary, we have shown the potential of using in situ digestion coupled with IMS for the high-throughput proteomic analysis of TMAs. The relative intensity and distribution of hundreds of proteins can be simultaneously mapped in a single experiment. We showed that these experiments are reproducible within and between TMAs and therefore the results from multiple analyses can be effectively compared. This platform provides a degree of speed and depth that cannot be achieved using any alternative technique, and therefore has the potential to revolutionize the molecular analysis of clinical samples.

Appendix A

Protein Identification

Search Parameters
Type of search: MS/MS Ion Search
Enzyme: Trypsin
Variable modifications: N-Acetyl, Ox. (M/H/W)
Mass values: Monoisotopic
Protein Mass: Unrestricted
Peptide Mass Tolerance : ± 200 ppm
Fragment Mass Tolerance: ± 0.4 Da
Max Missed Cleavages: 3
Instrument type: MALDI-TOF-TOF
Number of queries: 123

	SwissProt	Decoy	False discovery rate
Matches above identity threshold	94	0	0
Matches above homology or identity threshold	102	1	0.98

Keratin, type II cytoskeletal 6A	Score (233)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1016.5420	1015.5347	1015.5298	4.83	0	41	0.0094	R.QLDSIVGER.G
1165.5830	1164.5757	1164.5775	-1.51	0	62	5.7e-05	K.YEELQVTAGR.H
1263.6915	1262.6842	1262.687	-2.22	0	81	9.2e-07	K.LALDVEIATYR.K
1407.7100	1406.7027	1406.7041	-1.01	0	34	0.044	K.ADTLTDEINFLR.A
1890.9550	1889.9477	1889.9635	-8.36	0	92	7.1e-08	R.QNLEPLFEQYINNLR.R
2047.0294	2046.0221	2046.0646	-20.77	1	21	0.88	R.QNLEPLFEQYINNLRR.Q
2471.1831	2470.1758	2470.1798	-1.61	3	82	7.1e-07	R.GMQDLVEDFKNKYEDEINKR.T

Keratin, type II cytoskeletal 5	Score (198)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1016.5420	1015.5347	1015.5298	4.83	0	41	0.0094	R.QLDSIVGER.G
1093.5220	1092.5147	1092.5199	-4.78	0	21	0.67	K.AQYEEIANR.S
1263.6915	1262.6842	1262.687	-2.22	0	81	9.2e-07	K.LALDVEIATYR.K
1410.7325	1409.7252	1409.7151	7.21	0	44	0.0042	R.SFSTASAITPSVSR.T
1439.7126	1438.7053	1438.7053	0.04	0	66	2.6e-05	R.GLGVGFGSGGGSSSSVK.F
1890.9550	1889.9477	1889.9635	-8.36	0	92	7.1e-08	R.QNLEPLFEQYINNLR.R
2047.0294	2046.0221	2046.0646	-20.77	1	21	0.88	R.QNLEPLFEQYINNLRR.Q

Keratin, type II cytoskeletal 8	Score (144)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1129.6152	1128.6079	1128.6138	-5.25	0	48	0.0019	K.LSELEAALQR.A
1277.6904	1276.6831	1276.7027	-15.31	0	22	0.7	K.LALDIEIATYR.K
1419.7931	1418.7858	1418.7405	31.9	0	64	4.7e-05	R.LEGLTDEINFLR.Q
1847.8475	1846.8402	1846.7978	23.0	0	83	5.8e-07	R.SNMDNMFESYINNLR.R
1863.8770	1862.8697	1862.7927	41.4	0	(49)	0.0013	R.SNMDNMFESYINNLR.R + Oxidation (M)

Histone H2A	Score (146)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
944.5400	943.5327	943.524	9.27	0	17	2.1	R.AGLQFPVGR.V
2915.5000	2914.4927	2914.5804	-30.07	0	146	2.8e-13	R.VGAGAPVYLAAVLEYLTAEILELAGNAAR.D

Alpha-enolase	Score (122)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1129.6152	1128.6079	1128.6138	-5.25	0	48	0.0019	K.LSELEAALQR.A
1277.6904	1276.6831	1276.7027	-15.31	0	22	0.7	K.LALDIEIATYR.K
1419.7931	1418.7858	1418.7405	31.9	0	64	4.7e-05	R.LEGLTDEINFLR.Q
1847.8475	1846.8402	1846.7978	23.0	0	83	5.8e-07	R.SNMDNMFESYINNLR.R
1863.8770	1862.8697	1862.7927	41.4	0	(49)	0.0013	R.SNMDNMFESYINNLR.R + Oxidation (M)

Keratin, type I cytoskeletal 19	Score (105)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1082.5175	1081.5102	1081.4829	25.3	0	27	0.16	K.DAEAWFTSR.T
1122.5560	1121.5487	1121.5717	-20.45	0	40	0.01	R.LEQEIATYR.S
1210.6031	1209.5958	1209.5778	14.9	1	32	0.06	R.KDAEAWFTSR.T
1222.6390	1221.6317	1221.6353	-2.95	1	55	0.00033	R.TKFETEQALR.M
1354.6100	1353.6027	1353.5983	3.28	0	37	0.018	R.SQYEVMAEQNR.K
1554.7480	1553.7407	1553.7434	-1.74	0	42	0.0065	R.QSSATSSFGGLGGGSVR.F
1674.8083	1673.8010	1673.7685	19.4	0	50	0.001	R.DYSHYYTTIQDLR.D
2407.2258	2406.2185	2406.3019	-34.66	0	30	0.12	R.FGAQLAHIQALISGIEAQLGDVR.A

Tubulin beta-2C chain	Score (118)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1130.6021	1129.5948	1129.588	6.04	0	64	4.2e-05	R.FPGQLNADLR.K
1143.6100	1142.6027	1142.627	-21.27	0	29	0.15	K.LAVNMVPFPR.L
1601.8101	1600.8028	1600.8131	-6.40	0	75	3.8e-06	R.AVLVDLEPGTMDSVR.S
1620.8260	1619.8187	1619.8283	-5.89	0	31	0.086	R.LHFFMPGFAPLTSR.G
2798.4248	2797.4175	2797.3361	29.1	0	37	0.023	R.SGPFGQIFRPDNFVFGQSGAGNNWAK.G

Serum albumin precursor	Score (113)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
927.4923	926.4850	926.4861	-1.18	0	24	0.34	K.YLYEIAR.R
1467.8109	1466.8036	1466.8358	-21.92	1	33	0.051	R.RHPDYSVVLLLR.L
1511.7955	1510.7882	1510.8355	-31.31	0	55	0.00038	K.VPQVSTPTLVEVSR.N
1623.7710	1622.7637	1622.7803	-10.21	0	59	0.00014	K.DVFLGMFLYEYAR.R
1639.7600	1638.7527	1638.7752	-13.72	0	(47)	0.0021	K.DVFLGMFLYEYAR.R + Oxidation (M)
1898.9000	1897.8927	1897.9879	-50.15	1	30	0.11	R.RHPYFYAPELLFFAK.R

Keratin, type II cytoskeletal 75	Score (110)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1165.5830	1164.5757	1164.5775	-1.51	0	62	5.7e-05	K.YEELQVTAGR.H
1263.6915	1262.6842	1262.6870	-2.22	0	81	9.2e-07	K.LALDVEIATYR.K

Keratin, type II cytoskeletal 2 epidermal	Score (103)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1263.6915	1262.6842	1262.687	-2.22	0	81	9.2e-07	K.LALDVEIATYR.K
1320.6698	1319.6625	1319.5756	65.9	0	53	0.00057	R.HGGGGGGFGGGGFGSR.S
1475.7600	1474.7527	1474.778	-17.13	0	21	0.83	R.FLEQQNQVLQTK.W

Keratin, type I cytoskeletal 16	Score (94)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1036.5272	1035.5199	1035.525	-4.90	1	29	0.11	K.IRDWYQR.Q
1122.5560	1121.5487	1121.5717	-20.45	0	40	0.01	R.LEQEIATYR.R
1241.5970	1240.5897	1240.587	2.22	0	35	0.031	K.NHEEEMLALR.G
1259.6740	1258.6667	1258.6669	-0.18	1	41	0.0082	R.TKYEHELALR.Q
1757.8680	1756.8607	1756.8784	-10.06	1	34	0.045	R.QRPSEIKDYSPYFK.T
2064.1130	2063.1057	2063.1375	-15.38	0	67	2.4e-05	K.IIAATIENAQPILQIDNAR.L

Vimentin	Score (88)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1428.7123	1427.7050	1427.7045	0.38	0	77	2.4e-06	R.SLYASSPGGVYATR.S
2497.2173	2496.2100	2496.2496	-15.86	1	41	0.0089	R.LLQDSVDFSLADAINTEFKNTR.T

Keratin, type I cytoskeletal 15	Score (108)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1122.5560	1121.5487	1121.5717	-20.45	0	40	0.01	R.LEQEIATYR.S
1248.6610	1247.6537	1247.6873	-26.94	1	45	0.0037	R.LKYENELALR.Q
1821.8787	1820.8714	1820.8766	-2.84	0	77	2.4e-06	R.GGSLLAGGGGFGGGSLSGGGGSR.S
1877.8959	1876.8886	1876.8803	4.42	0	41	0.0079	M.TTTFLQTSSSTFGGGSTR.G + Acetyl (N-term)

Ferritin light chain	Score (101)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1607.7820	1606.7747	1606.7991	-15.17	0	84	4.5e-07	R.LGGPEAGLGEYLFER.L
1719.9220	1718.9147	1718.9791	-37.46	1	49	0.0014	K.KLNQALLDLHALGSAR.T

Protein S100-A9	Score (72)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1455.7142	1454.7069	1454.7154	-5.82	0	18	1.6	K.LGHPDTLNQGEFK.E
1614.7900	1613.7827	1613.7945	-7.32	0	53	0.00062	K.QLSFEEFIMLMAR.L
1806.9113	1805.9040	1805.9312	-15.04	0	51	0.00087	R.NIETIINTFHQYSVK.L

Keratin, type I cytoskeletal 13	Score (71)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1122.5560	1121.5487	1121.5717	-20.45	0	40	0.01	R.LEQEIATYR.S
1248.6610	1247.6537	1247.6873	-26.94	1	45	0.0037	R.LKYENELALR.Q
1392.6880	1391.6807	1391.6867	-4.32	0	51	0.00082	K.MIGFPSSAGSVSPR.S

78 kDa glucose-regulated protein precursor	Score (78)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1566.7791	1565.7718	1565.7726	-0.48	0	51	0.00078	R.ITPSYVAFTPEGER.L
1999.0770	1998.0697	1998.0786	-4.44	0	58	0.0002	R.GVPQIEVTFEIDVNGILR.

Keratin, type I cytoskeletal 17	Score (62)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1036.5272	1035.5199	1035.5250	-4.90	1	29	0.11	K.IRDWYQR.Q
1122.5560	1121.5487	1121.5717	-20.45	0	40	0.01	R.LEQEIATYR.R
1222.6390	1221.6317	1221.6353	-2.95	1	55	0.00033	R.TKFETEQALR.L
2114.0249	2113.0176	2113.011	3.14	0	26	0.32	R.GQVGGEINVEMDAAPGVDLSR.I

119

Pulmonary surfactant-associated protein A1 precursor	Score (103)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1447.7350	1446.7277	1446.699	19.8	0	52	0.00069	R.NPEENEAIASFVK.K
1506.8000	1505.7927	1505.7759	11.1	0	23	0.51	R.GALSLQGSIMTVGEK.V + Ox.(M)
1984.0176	1983.0103	1983.0425	-16.22	1	37	0.023	R.IAVPRNPEENEAIASFVK.K
2106.9856	2105.9783	2105.9694	4.22	0	48	0.002	K.YNTYAYVGLTEGPSPGDFR.Y
2329.1340	2328.1267	2328.1499	-9.94	0	61	8.4e-05	R.GPPGLPAHLDEELQATLHDFR.H + Ox. (HW)

Tubulin beta-2A chain	Score (74)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1130.6021	1129.5948	1129.588	6.04	0	64	4.2e-05	R.FPGQLNADLR.K
1143.6100	1142.6027	1142.627	-21.27	0	29	0.15	K.LAVNMVPFPR.L
1615.8186	1614.8113	1614.8287	-10.76	0	20	1.1	R.AILVDLEPGTMDSVR.S
1620.8260	1619.8187	1619.8283	-5.89	0	31	0.086	R.LHFFMPGFAPLTSR.G
2798.4248	2797.4175	2797.3361	29.1	0	37	0.023	R.SGPFGQIFRPDNFVFGQSGAGNNWAK.G

Histone H4	Score (57)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
989.5700	988.5627	988.5706	-7.91	0	36	0.025	K.VFLENVIR.D
1325.7233	1324.7160	1324.7463	-22.84	0	54	0.00044	R.DNIQGITKPAIR.R

Keratin, type I cytoskeletal 18	Score (82)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
965.4800	964.4727	964.4614	11.7	0	36	0.024	R.AQYDELAR.K
975.4492	974.4419	974.4458	-3.91	0	31	0.085	R.STFSTNYR.S
2670.3108	2669.3035	2669.3846	-30.38	0	48	0.0018	R.YALQMEQLNGILLHLESELAQTR.A
2854.3882	2853.3809	2853.4005	-6.88	0	50	0.0012	R.SLGSVQAPSYGARPVSSAASVYAGAGGSGSR.I

Actin, cytoplasmic	Score (90)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
976.4540	975.4467	975.441	5.84	0	30	0.1	K.AGFAGDDAPR.A
1515.6600	1514.6527	1514.7419	-58.85	0	59	0.00014	K.IWHHTFYNELR.V
1790.8907	1789.8834	1789.8846	-0.67	0	61	8.9e-05	K.SYELPDGQVITIGNER.F
2215.0000	2213.9927	2214.0627	-31.60	0	28	0.19	K.DLYANTVLSGGTTMYPGIADR.M

Vitronectin precursor	Score (56)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1314.6980	1313.6907	1313.6728	13.6	1	29	0.14	R.RVDTVDPPYPR.S
1646.8475	1645.8402	1645.81	18.3	0	54	0.00042	R.DVWGIEGPIDAAFTR.I
1666.7800	1665.7727	1665.7682	2.72	0	30	0.11	R.DWHGVPGQVDAAMAGR.I

Glutathione S-transferase P	Score (46)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1337.7000	1336.6927	1336.718	-18.89	0	29	0.13	M.PPYTVVYFPVR.G
1883.9390	1882.9317	1882.9425	-5.72	0	46	0.0031	K.FQDGDLTLYQSNTILR.H

Keratin, type II cytoskeletal 7	Score (61)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1277.6904	1276.6831	1276.7027	-15.31	0	22	0.7	K.LALDIEIATYR.K
1406.7100	1405.7027	1405.699	2.63	0	61	7.9e-05	M.SIHFSSPVFTSR.S + Acetyl (N-term)
1442.7550	1441.7477	1441.7929	-31.33	0	21	0.89	R.LPDIFEAQIAGLR.G
1453.8199	1452.8126	1452.83	-11.96	0	28	0.16	R.EVTINQSLLAPLR.L

Ig alpha chain C region	Score (63)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1213.6259	1212.6186	1212.6251	-5.35	0	32	0.064	R.WLQGSQELPR.E
1835.9080	1834.9007	1834.9425	-22.77	0	63	5.4e-05	R.QEPSQGTTTFAVTSILR.V

Pyruvate kinase isozymes M1/M2	Score (68)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1859.8700	1858.8627	1858.8924	-15.96	0	46	0.0026	K.FGVEQDVDMVFASFIR.K
2465.2520	2464.2447	2464.2849	-16.32	0	55	0.00036	R.TATESFASDPILYRPVAVALDTK.G

Fibrinogen beta chain precursor	Score (50)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
980.4486	979.4413	979.4359	5.50	0	39	0.011	R.QDGSVDFGR.K
1239.5300	1238.5227	1238.5105	9.88	0	25	0.26	K.EDGGGWWYNR.C
1950.9750	1949.9677	1949.9959	-14.45	0	39	0.015	R.EEAPSLRPAPPPISGGGYR.A

Fibrinogen gamma chain precursor	Score (47)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1117.5070	1116.4997	1116.52	-18.14	0	38	0.019	R.VELEDWNGR.T
1682.9268	1681.9195	1681.9515	-19.02	0	40	0.0099	K.IHLISTQSAIPYALR.V

Heat shock protein beta-1	Score (50)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
987.5430	986.5358	986.6025	-67.68	1	42	0.0069	R.RVPFSLLR.G
1163.6035	1162.5962	1162.6135	-14.86	0	20	0.94	R.LFDQAFGLPR.L
1905.9900	1904.9827	1904.9843	-0.85	0	41	0.0095	K.LATQSNEITIPVTFESR.A

Proteins Identified with De Novo Sequencing

Collagen alpha-1(I) chain precursor					
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Peptide
1105.578	1104.570	1104.567	2.8	0	GVQGPPGPAGPR (Hydroxylated)
1459.711	1458.703	1458.759	38.4	0	GSAGPPGATGFPGAAGR (Hydroxylated)
898.484	897.476	897.502	29.5	0	GVVGLPGQR (Hydroxylated)
2705.260	2704.252	2704.247	1.9	0	GFSGLQGPPGPPGSPGEQGPSGASGPAGPR (Hydroxylated)
2869.418	2868.410	2868.399	3.7	1	GLTGPIGPPGPAGAPGDKGESGPSGPAGPTGAR (Hydroxylated)

Collagen alpha-2(I) chain precursor					
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Peptide
1477.760	1476.752	1476.747	3.7	0	GLHGEFGLPGPAGPR (Hydroxylated)
1562.804	1561.796	1561.781	9.9	0	GETGPSGPVGPAGAVGPR
1493.771	1492.763	1492.742	14.4	0	GLHGEFGLPGPAGPR (Hydroxylated)
1533.719	1532.711	1532.667	28.7	0	GDGGPPGMTGFPGAAGR (Hydroxylated)
1655.823	1654.815	1654.806	5.7	1	GFPGADGVAGPKGPAGER (Hydroxylated)
2959.415	2958.407	2958.348	19.8	0	GPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGAR (Hydroxylated)

Tentative Protein Identifications (<2 Peptides)

Protein disulfide-isomerase precursor	Score (50)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1780.8400	1779.8327	1779.8275	2.93	0	50	0.001	K.VDATEESDLAQQYGVR.G

Fibrinogen alpha chain precursor	Score (78)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1520.7173	1519.7100	1519.7267	-10.96	0	78	1.7e-06	K.GLIDEVNQDFTNR.I

Aldo-keto reductase family 1 member C1	Score (85)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
2179.0750	2178.0677	2178.0817	-6.44	0	85	3.8e-07	R.HIDSAHLYNNEEQVGLAIR.S

Glyceraldehyde-3-phosphate dehydrogenase	Score (79)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1763.8364	1762.8291	1762.7951	19.3	0	79	1.5e-06	K.LISWYDNEFGYSNR.V

Carbonyl reductase [NADPH] 1	Score (72)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1652.8300	1651.8227	1651.8642	-25.09	0	72	8e-06	R.GQAAVQQLQAEGLSPR.F

Hemoglobin subunit alpha	Score (69)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1833.9000	1832.8927	1832.8846	4.43	0	69	1.6e-05	K.TYFPHFDLSHGSAQVK.G

Annexin A1	Score (63)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1702.8875	1701.8802	1701.8785	1.03	0	63	6e-05	K.GLGTDEDTLIEILASR.T

Carcinoembryonic antigen-related cell adhesion molecule 5	Score (69)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
2177.0930	2176.0857	2176.1277	-19.28	0	69	1.4e-05	R.QIIGYVIGTQQATPGPAYSGR.E

Apolipoprotein A-I precursor	Score (53)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1301.6510	1300.6437	1300.6411	1.98	0	53	0.00059	R.THLAPYSDELR.Q

Histone H2B	Score (47)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1759.8075	1758.8002	1758.8069	-3.80	0	47	0.0015	K.AMGIMNSFVNDIFER.I + Ox. (M)

Annexin A2	Score (47)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1111.5660	1110.5587	1110.5458	11.6	0	47	0.0019	R.QDIAFAYQR.R

Macrophage migration inhibitory factor	Score (50)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1287.6816	1286.6743	1286.6805	-4.82	0	50	0.001	M.PMFIVNTNVPR.A

Serum amyloid P-component precursor	Score (45)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1156.5970	1155.5897	1155.5924	-2.32	0	45	0.0034	R.VGEYSLYIGR.H

Tubulin alpha-1A chain	Score (45)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1701.8616	1700.8543	1700.8985	-25.98	0	45	0.0037	R.AVFVDLEPTVIDEVR.T

Stathmin	Score (49)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1388.7150	1387.7077	1387.7459	-27.54	0	49	0.0012	R.ASGQAFELILSPR.S

Peptidyl-prolyl cis-trans isomerase A	Score (46)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1987.9899	1986.9826	1987.0051	-11.32	0	46	0.0029	M.VNPTVFFDIAVDGEPLGR.V + Acetyl (N-term)

Protein disulfide-isomerase A3 precursor	Score (46)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1370.7358	1369.7285	1369.6878	29.8	0	46	0.0024	R.ELSDFISYLQR.E

Histone H3-like	Score (45)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
788.4709	787.4636	787.4704	-8.67	1	45	0.0035	R.KLPFQR.L

Polymeric immunoglobulin receptor precursor	Score (48)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1228.7552	1227.7479	1227.7551	-5.82	0	48	0.0019	R.LVSLTLNLVTR.A

Heterogeneous nuclear ribonucleoproteins A2/B1	Score (46)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1377.6200	1376.6127	1376.6222	-6.88	0	46	0.0024	R.GGGGNFGPGPGSNFR.G

Protein S100-P	Score (47)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1854.8900	1853.8827	1853.8903	-4.09	0	47	0.0023	M.TELETAMGMIIDVFSR.Y + Acetyl (N-term)

Myosin-9 - Homo sapiens (Human)	Score (43)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1869.9500	1868.9427	1868.9592	-8.80	0	43	0.005	K.ANLQIDQINTDLNLER.S

Keratin, type II cytoskeletal 3	Score (81)						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1263.6915	1262.6842	1262.6870	-2.22	0	81	9.2e-07	K.LALDVEIATYR.K

Hemoglobin subunit beta	Score (66)						
1274.7000	1273.6927	1273.7183	-20.07	0	66	2.9e-05	R.LLVVYPWTQR.F

Fully Annotated MS/MS Spectra

Fibrinogen alpha chain precursor		Mascot Score (61)						
Observed		Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1520.7173		1519.7100	1519.7267	-10.96	0	61	5.4e-05	K.GLIDEVNQDFTNR.I

	Gly	Leu	Ile	Asp	Glu	Val	Asn	Gln	Asp	Phe	Thr	Asn	Arg
Ion	1	2	3	4	5	6	7	8	9	10	11	12	13
a													
b		0.043	0.02		0.043								
b-17													
b-18													
b+18													-0.017
y	0.032	0.011	0.058	0.048	0.056	-0.064	-0.017	-0.088	-0.019	-0.058			-0.017
y-17	0.042	0.023	0.026		0.004		0.002						
i		0	0										0.036
	Arg	Asn	Thr	Phe	Asp	Gln	Asn	Val	Glu	Asp	Ile	Leu	Gly

124

Apolipoprotein A-I precursor		Mascot Score (48)						
Observed		Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1301.6510		1300.6437	1300.6411	1.98	0	48	0.0013	R.THLAPYSDELR.Q

Fragment Ion Mass Errors (Da)											
	Thr	His	Leu	Ala	Pro	Tyr	Ser	Asp	Glu	Leu	Arg
Ion	1	2	3	4	5	6	7	8	9	10	11
a		0.036									
b		0.009	0.043	0.066					0.053		
b-17										-0.539	
b-18	0.043									0.445	
b+18											0.015
y	0.022	0.001	0.047	0.031	-0.017		0.043	0.159	0.042		0.015
y-17		-0.023		0.007	0.047		0.102				
i		0.021	0.057		0.01	0.014				0.057	0.012
	Arg	Leu	Glu	Asp	Ser	Tyr	Pro	Ala	Leu	His	Thr

125

Serum amyloid P-component precursor		Mascot Score (43)						
Observed		Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1156.5970		1155.5897	1155.592	-2.32	0	43	0.0041	R.VGEYSLYIGR.H

Fragment Ion Mass Errors (Da)

	Val	Gly	Glu	Tyr	Ser	Leu	Tyr	Ile	Gly	Arg
Ion	1	2	3	4	5	6	7	8	9	10
a		0.014			0.082	0.002				
b			0.009	0.037	0.035	-0.025				
b-17										
b-18										
b+18		0.011								-0.003
y	0	0.012	-0.018	0.035	-0.046	0.016	0.085			-0.003
y-17	-0.015	-0.039	-0.009	-0.056						
i						-0.01		-0.01		0.003
	Arg	Gly	Ile	Tyr	Leu	Ser	Tyr	Glu	Gly	Val

Histone H3-like	Mascot Score37						
Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
788.4709	787.4636	787.4704	-8.67	1	37	0.019	R.KLPFQR.L

Fragment Ion Mass Errors (Da)						
	Lys	Leu	Pro	Phe	Gln	Arg
Ion	1	2	3	4	5	6
a	-0.015					
b	0.011	0.02		-0.114		
b-17	0.019		0.002			
b-18						
b+18						0.044
y	0.01	0.032	-0.026	0.014	0.021	0.044
y-17	-0.019	0.053	0.034	-0.016	0.004	
i	-0.015	-0.001	-0.007		0.022	0
	Arg	Gln	Phe	Pro	Leu	Lys

127

Annexin A2		Mascot Score (46)						
Observed		Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Peptide
1111.5510		1110.5437	1110.5458	-1.86	0	46	0.0015	R.QDIAFAYQR.R

Fragment Ion Mass Errors (Da)									
	Gln	Asp	Ile	Ala	Phe	Ala	Tyr	Gln	Arg
Ion	1	2	3	4	5	6	7	8	9
a	0.019		0.015			0.012			
b	0.028	0.037	0	0.042	0.049				
b-17	0.058		0.034	0.041	0.002				
b-18		0.013		0.026					
b+18									0.013
y	-0.014	0.029	0.011		-0.016	0.209	0.081	0.108	0.013
y-17	-0.012	-0.013	0.012		0.058				
i	0.019		0		0.015		0.02	0.019	-0.019
	Arg	Gln	Tyr	Ala	Phe	Ala	Ile	Asp	Gln

REFERENCES

1.   Atkinson Jr, A. J.; Colburn, W. A.; DeGruttola, V. G.; DeMets, D. L.; Downing, G. J.; Hoth, D. F.; Oates, J. A.; Peck, C. C.; Schooley, R. T.; Spilker, B. A.; Woodcock, J.; Zeger, S. L., Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clincial Pharmacology & Therapeutics* **2001,** *69*, 89-95.

2.   Zhang, H.; Chan, D. W., Cancer biomarker discovery in plasma using a tissue-targeted proteomic approach. *Cancer Epidemiol Biomarkers Prev* **2007,** *16* (10), 1915-7.

3.   Simpson, R. J.; Bernhard, O. K.; Greening, D. W.; Moritz, R. L., Proteomics-driven cancer biomarker discovery: looking to the future. *Curr Opin Chem Biol* **2008,** *12* (1), 72-7.

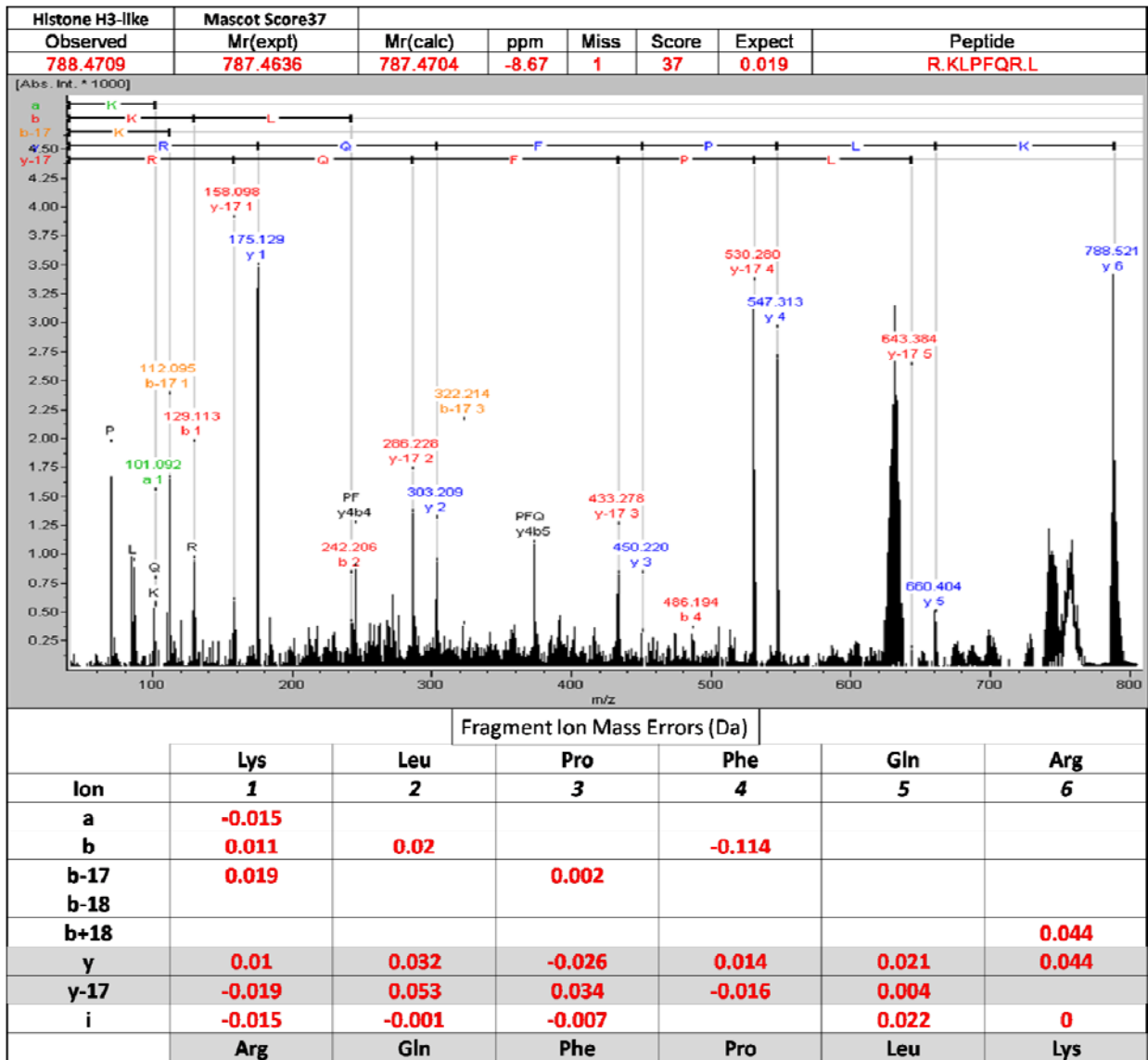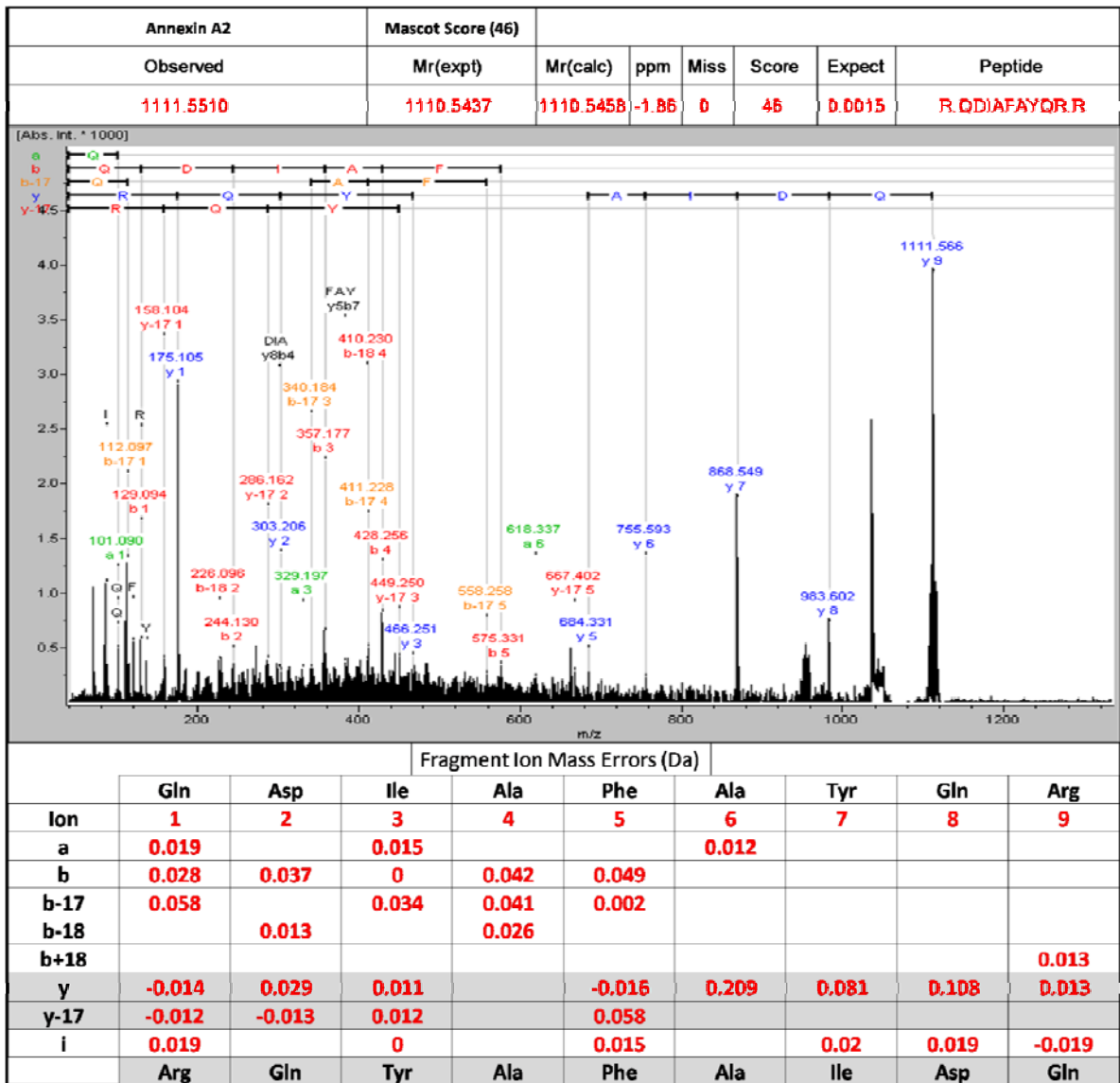4.   Etzioni, R.; Urban, N.; Ramsey, S.; McIntosh, M.; Schwartz, S.; Reid, B.; Radich, J.; Anderson, G.; Hartwell, L., The case for early detection. *Nat Rev Cancer* **2003,** *3* (4), 243-52.

5.   Nolen, B. M.; Marks, J. R.; Ta'san, S.; Rand, A.; Luong, T. M.; Wang, Y.; Blackwell, K.; Lokshin, A. E., Serum biomarker profiles and response to neoadjuvant chemotherapy for locally advanced breast cancer. *Breast Cancer Res* **2008,** *10* (3), R45.

6.   Vasudev, N. S.; Ferguson, R. E.; Cairns, D. A.; Stanley, A. J.; Selby, P. J.; Banks, R. E., Serum biomarker discovery in renal cancer using 2-DE and prefractionation by immunodepletion and isoelectric focusing; increasing coverage or more of the same? *Proteomics* **2008,** *8* (23-24), 5074-85.

7.   Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **2002,** *1* (11), 845-67.

8.   Weissleder, R.; Mahmood, U., Molecular imaging. *Radiology* **2001,** *219* (2), 316-33.

9.   Yuste, R., Fluorescence microscopy today. *Nat Methods* **2005,** *2* (12), 902-4.

10.   Sullivan, C. A.; Chung, G. G., Biomarker validation: in situ analysis of protein expression using semiquantitative immunohistochemistry-based techniques. *Clin Colorectal Cancer* **2008,** *7* (3), 172-7.

11.   Van den Abbeele, A. D., The lessons of GIST--PET and PET/CT: a new paradigm for imaging. *Oncologist* **2008,** *13 Suppl 2*, 8-13.

12.   Martin, W. R., MR spectroscopy in neurodegenerative disease. *Mol Imaging Biol* **2007,** *9* (4), 196-203.

13.   Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M., Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nature medicine* **2001,** *7* (4), 493-6.

14.   Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M., MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat Methods* **2007,** *4* (10), 828-33.

15.   McDonnell, L. A.; Heeren, R. M. A., Imaging mass spectrometry. *Mass Spectrometry Reviews* **2007,** *26* (4), 606-643.

16.   Burns, M. S., Applications of secondary ion mass spectrometry (SIMS) in biological research: a review. *J Microsc* **1982,** *127* (Pt 3), 237-58.

17.   Brunelle, A.; Laprevote, O., Recent advances in biological tissue imaging with Time-of-flight Secondary Ion Mass Spectrometry: polyatomic ion sources, sample preparation, and applications. *Curr Pharm Des* **2007,** *13* (32), 3335-43.

18.   Touboul, D.; Kollmer, F.; Niehuis, E.; Brunelle, A.; Laprevote, O., Improvement of biological time-of-flight-secondary ion mass spectrometry imaging with a bismuth cluster ion source. *J Am Soc Mass Spectrom* **2005,** *16* (10), 1608-18.

19.   Brunelle, A.; Touboul, D.; Laprevote, O., Biological tissue imaging with time-of-flight secondary ion mass spectrometry and cluster ion sources. *J Mass Spectrom* **2005,** *40* (8), 985-99.
20.   Mas, S.; Perez, R.; Martinez-Pinna, R.; Egido, J.; Vivanco, F., Cluster TOF-SIMS imaging: a new light for in situ metabolomics? *Proteomics* **2008,** *8* (18), 3735-45.

21.   Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988,** *60* (20), 2299-301.

22.   Koichi Tanaka, H. W., Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida, Tamio Yoshida, T. Matsuo,, Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **1988,** *2* (8), 151-153.

23.   Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

24.   Caprioli, R. M.; Farmer, T. B.; Gile, J., Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal Chem* **1997,** *69* (23), 4751-60.

25.   Seeley, E. H.; Caprioli, R. M., Molecular imaging of proteins in tissues by mass spectrometry. *Proc Natl Acad Sci U S A* **2008,** *105* (47), 18126-31.

26.   McDonnell, L. A.; Heeren, R. M., Imaging mass spectrometry. *Mass Spectrom Rev* **2007,** *26* (4), 606-43.

27.   Schwartz, S. A.; Reyzer, M. L.; Caprioli, R. M., Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *J Mass Spectrom* **2003,** *38* (7), 699-708.

28.   Chaurand, P.; Schwartz, S. A.; Capriolo, R. M., Profiling and imaging proteins in tissue sections by MS. *Anal Chem* **2004,** *76* (5), 87A-93A.

29.   Seeley, E. H.; Oppenheimer, S. R.; Mi, D.; Chaurand, P.; Caprioli, R. M., Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. *J Am Soc Mass Spectrom* **2008,** *19* (8), 1069-77.

30.   Murphy, R. C.; Hankin, J. A.; Barkley, R. M., Imaging of lipid species by MALDI mass spectrometry. *J Lipid Res* **2008**.

31.   Jaskolla, T. W.; Lehmann, W. D.; Karas, M., 4-Chloro-alpha-cyanocinnamic acid is an advanced, rationally designed MALDI matrix. *Proc Natl Acad Sci U S A* **2008,** *105* (34), 12200-5.

32.   Krause, E.; Wenschuh, H.; Jungblut, P. R., The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal Chem* **1999,** *71* (19), 4160-5.

33.   Peterson, D. S., Matrix-free methods for laser desorption/ionization mass spectrometry. *Mass Spectrom Rev* **2007,** *26* (1), 19-34.

34.   Tholey, A.; Heinzle, E., Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry-applications and perspectives. *Anal Bioanal Chem* **2006,** *386* (1), 24-37.

35.   Drexler, D. M.; Garrett, T. J.; Cantone, J. L.; Diters, R. W.; Mitroka, J. G.; Prieto Conaway, M. C.; Adams, S. P.; Yost, R. A.; Sanders, M., Utility of imaging mass spectrometry (IMS) by matrix-assisted laser desorption ionization (MALDI) on an ion trap mass spectrometer in the analysis of drugs and metabolites in biological tissues. *J Pharmacol Toxicol Methods* **2007,** *55* (3), 279-88.

36.   Dill, A. L.; Ifa, D. R.; Manicke, N. E.; Ouyang, Z.; Cooks, R. G., Mass spectrometric imaging of lipids using desorption electrospray ionization. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**.

37.   Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F., Matrix-Assisted Ultraviolet-Laser Desorption of Nonvolatile Compounds. *International Journal of Mass Spectrometry and Ion Processes* **1987,** *78*, 53-68.

38.   Wisniewski, J. R., Mass spectrometry-based proteomics: principles, perspectives, and challenges. *Arch Pathol Lab Med* **2008,** *132* (10), 1566-9.

39.   Chen, W. Y.; Chen, Y. C., MALDI MS analysis of oligonucleotides: desalting by functional magnetite beads using microwave-assisted extraction. *Anal Chem* **2007,** *79* (21), 8061-6.

40.   Fu, Y.; Xu, S.; Pan, C.; Ye, M.; Zou, H., 3,4-diaminobenzophenone matrix for analysis of oligonucleotides by MALDI-TOF mass spectrometry. *Curr Protoc Nucleic Acid Chem* **2007,** *Chapter 10*, Unit 10 12.

41.   Harvey, D. J., Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 1999-2000. *Mass Spectrom Rev* **2006,** *25* (4), 595-662.

42.   Irungu, J.; Go, E. P.; Zhang, Y.; Dalpathado, D. S.; Liao, H. X.; Haynes, B. F.; Desaire, H., Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J Am Soc Mass Spectrom* **2008,** *19* (8), 1209-20.

43.   Yu, Y. Q.; Fournier, J.; Gilar, M.; Gebler, J. C., Identification of N-linked glycosylation sites using glycoprotein digestion with pronase prior to MALDI tandem time-of-flight mass spectrometry. *Anal Chem* **2007,** *79* (4), 1731-8.

44.   Snovida, S. I.; Chen, V. C.; Krokhin, O.; Perreault, H., Isolation and identification of sialylated glycopeptides from bovine alpha1-acid glycoprotein by off-line capillary electrophoresis MALDI-TOF mass spectrometry. *Anal Chem* **2006,** *78* (18), 6556-63.

45.   Trimpin, S.; Weidner, S. M.; Falkenhagen, J.; McEwen, C. N., Fractionation and solvent-free MALDI-MS analysis of polymers using liquid adsorption chromatography at critical conditions in combination with a multisample on-target homogenization/transfer sample preparation method. *Anal Chem* **2007,** *79* (19), 7565-70.

46.   Wetzel, S. J.; Guttman, C. M.; Flynn, K. M.; Filliben, J. J., Significant parameters in the optimization of MALDI-TOF-MS for synthetic polymers. *J Am Soc Mass Spectrom* **2006,** *17* (2), 246-52.

47.   Gluckmann, M.; Pfenninger, A.; Kruger, R.; Thierolf, M.; Karas, M.; Horneffer, V.; Hillenkamp, F.; Strupat, K., Mechanisms in MALDI analysis: surface interaction or incorporation of analytes? *International Journal of Mass Spectrometry* **2001,** *210* (1-3), 121-132.

48.   Chang, W. C.; Huang, L. C. L.; Wang, Y. S.; Peng, W. P.; Chang, H. C.; Hsu, N. Y.; Yang, W. B.; Chen, C. H., Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. *Analytica Chimica Acta* **2007,** *582* (1), 1-9.

49.   Karas, M.; Kruger, R., Ion formation in MALDI: The cluster ionization mechanism. *Chemical Reviews* **2003,** *103* (2), 427-439.

50.   Knochenmuss, R., Ion formation mechanisms in UV-MALDI. *Analyst* **2006,** *131* (9), 966-986.

51.   Whittal, R. M.; Li, L., High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. *Anal Chem* **1995,** *67* (13), 1950-4.

52.   M. L. Vestal · P. J., S. A. Martin, Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **1995,** *9* (11), 1044-1050.

53.   Brown, R. S.; Lennon, J. J., Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Anal Chem* **1995,** *67* (13), 1998-2003.

54.   Mamyrin, B. A., Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry* **2001,** *206*, 251-266.

55.   Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A., A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and Bioanalytical Chemistry* **2003,** *376* (7), 952-965.

56.   Kaufmann, R.; Kirsch, D.; Spengler, B., Sequencing of Peptides in a Time-of-Flight Mass-Spectrometer - Evaluation of Postsource Decay Following Matrix-Assisted Laser-Desorption Ionization (Maldi). *International Journal of Mass Spectrometry and Ion Processes* **1994,** *131*, 355-385.

57.   Chaurand, P.; Schwartz, S. A.; Billheimer, D.; Xu, B. J.; Crecelius, A.; Caprioli, R. M., Integrating histology and imaging mass spectrometry. *Anal Chem* **2004,** *76* (4), 1145-55.

58.   Schwartz, S. A.; Weil, R. J.; Johnson, M. D.; Toms, S. A.; Caprioli, R. M., Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. *Clin Cancer Res* **2004,** *10* (3), 981-7.

59.   Palmer-Toy, D. E.; Sarracino, D. A.; Sgroi, D.; LeVangie, R.; Leopold, P. E., Direct acquisition of matrix-assisted laser Desorption/Ionization time-of-flight mass spectra from laser capture microdissected tissues. *Clin Chem* **2000,** *46* (9), 1513-6.

60.   Xu, B. J.; Caprioli, R. M.; Sanders, M. E.; Jensen, R. A., Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J Am Soc Mass Spectrom* **2002,** *13* (11), 1292-7.

61.   Masumori, N.; Thomas, T. Z.; Chaurand, P.; Case, T.; Paul, M.; Kasper, S.; Caprioli, R. M.; Tsukamoto, T.; Shappell, S. B.; Matusik, R. J., A probasin-large T antigen transgenic mouse line develops prostate adenocarcinoma and neuroendocrine carcinoma with metastatic potential. *Cancer Res* **2001,** *61* (5), 2239-49.

62.   Schwamborn, K.; Krieg, R. C.; Reska, M.; Jakse, G.; Knuechel, R.; Wellmann, A., Identifying prostate carcinoma by MALDI-Imaging. *Int J Mol Med* **2007,** *20* (2), 155-9.

63.   Chaurand, P.; DaGue, B. B.; Pearsall, R. S.; Threadgill, D. W.; Caprioli, R. M., Profiling proteins from azoxymethane-induced colon tumors at the molecular level by matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* **2001,** *1* (10), 1320-6.

64.   Bhattacharya, S. H.; Gal, A. A.; Murray, K. K., Laser capture microdissection MALDI for direct analysis of archival tissue. *J Proteome Res* **2003,** *2* (1), 95-8.

65. Yanagisawa, K.; Shyr, Y.; Xu, B. J.; Massion, P. P.; Larsen, P. H.; White, B. C.; Roberts, J. R.; Edgerton, M.; Gonzalez, A.; Nadaf, S.; Moore, J. H.; Caprioli, R. M.; Carbone, D. P., Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* **2003,** *362* (9382), 433-9.

66. Caldwell, R. L., Gonzalez, A., Oppenheimer, S. R., Schwartz,; H. S., C., R. M., Molecular assessment of the tumor protein microenvironment using imaging mass spectrometry *Cancer Genomics Proteomics* **2006,** *3*, 279-288.

67. Cornett, D. S.; Mobley, J. A.; Dias, E. C.; Andersson, M.; Arteaga, C. L.; Sanders, M. E.; Caprioli, R. M., A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Mol. Cell. Proteomics* **2006,** *5* (10), 1975-1983.

68. Rahman, S. M.; Shyr, Y.; Yildiz, P. B.; Gonzalez, A. L.; Li, H.; Zhang, X.; Chaurand, P.; Yanagisawa, K.; Slovis, B. S.; Miller, R. F.; Ninan, M.; Miller, Y. E.; Franklin, W. A.; Caprioli, R. M.; Carbone, D. P.; Massion, P. P., Proteomic patterns of preinvasive bronchial lesions. *Am J Respir Crit Care Med* **2005,** *172* (12), 1556-62.

69. Schwartz, S. A.; Weil, R. J.; Thompson, R. C.; Shyr, Y.; Moore, J. H.; Toms, S. A.; Johnson, M. D.; Caprioli, R. M., Proteomic-Based Prognosis of Brain Tumor Patients Using Direct-Tissue Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Cancer Res* **2005,** *65* (17), 7674-7681.

70. Bienvenut, W. V.; Sanchez, J. C.; Karmime, A.; Rouge, V.; Rose, K.; Binz, P. A.; Hochstrasser, D. F., Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. *Anal Chem* **1999,** *71* (21), 4800-7.

71. Binz, P. A.; Muller, M.; Hoogland, C.; Zimmermann, C.; Pasquarello, C.; Corthals, G.; Sanchez, J. C.; Hochstrasser, D. F.; Appel, R. D., The molecular scanner: concept and developments. *Curr Opin Biotechnol* **2004,** *15* (1), 17-23.

72. Muller, M.; Gras, R.; Appel, R. D.; Bienvenut, W. V.; Hochstrasser, D. F., Visualization and analysis of molecular scanner peptide mass spectra. *J Am Soc Mass Spectrom* **2002,** *13* (3), 221-31.

73. Jespersen, S.; Chaurand, P.; van Strien, F. J.; Spengler, B.; van der Greef, J., Direct sequencing of neuropeptides in biological tissue by MALDI-PSD mass spectrometry. *Anal Chem* **1999,** *71* (3), 660-6.

74. Jimenez, C. R.; Li, K. W.; Dreisewerd, K.; Spijker, S.; Kingston, R.; Bateman, R. H.; Burlingame, A. L.; Smit, A. B.; van Minnen, J.; Geraerts, W. P., Direct mass spectrometric peptide profiling and sequencing of single neurons reveals differential peptide patterns in a small neuronal network. *Biochemistry* **1998,** *37* (7), 2070-6.

75. Kruse, R.; Sweedler, J. V., Spatial profiling invertebrate ganglia using MALDI MS. *J Am Soc Mass Spectrom* **2003,** *14* (7), 752-9.

76.   Kutz, K. K.; Schmidt, J. J.; Li, L., In situ tissue analysis of neuropeptides by MALDI FTMS in-cell accumulation. *Anal Chem* **2004,** *76* (19), 5630-40.

77.   Li, L.; Garden, R. W.; Romanova, E. V.; Sweedler, J. V., In situ sequencing of peptides from biological tissues and single cells using MALDI-PSD/CID analysis. *Anal Chem* **1999,** *71* (24), 5451-8.

78.   Reyzer, M. L.; Hsieh, Y.; Ng, K.; Korfmacher, W. A.; Caprioli, R. M., Direct analysis of drug candidates in tissue by matrix-assisted laser desorption/ionization mass spectrometry. *J Mass Spectrom* **2003,** *38* (10), 1081-92.

79.   Rubakhin, S. S.; Greenough, W. T.; Sweedler, J. V., Spatial profiling with MALDI MS: distribution of neuropeptides within single neurons. *Anal Chem* **2003,** *75* (20), 5374-80.

80.   Troendle, F. J.; Reddick, C. D.; Yost, R. A., Detection of pharmaceutical compounds in tissue by matrix-assisted laser desorption/ionization and laser desorption/chemical ionization tandem mass spectrometry with a quadrupole ion trap. *Journal of the American Society for Mass Spectrometry* **1999,** *10* (12), 1315-1321.

81.   Khatib-Shahidi, S.; Andersson, M.; Herman, J. L.; Gillespie, T. A.; Caprioli, R. M., Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Anal Chem* **2006,** *78* (18), 6448-56.

82.   Shimma, S.; Furuta, M.; Ichimura, K.; Yoshida, Y.; Setou, M., A Novel Approach to *in situ* Proteome Analysis Using Chemical Inkjet Printing Technology and MALDI-QIT-TOF Tandem Mass Spectrometer. *Journal of the Mass Spectrometry Society of Japan* **2006,** *54* (4), 133-140.

83.   Groseclose, M. R.; Andersson, M.; Hardesty, W. M.; Caprioli, R. M., Identification of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry. *J Mass Spectrom* **2007,** *42* (2), 254-62.

84.   Aerni, H. R.; Cornett, D. S.; Caprioli, R. M., Automated acoustic matrix deposition for MALDI sample preparation. *Anal Chem* **2006,** *78* (3), 827-34.

85.   Boggs, J. M., Myelin basic protein: a multifunctional protein. *Cell Mol Life Sci* **2006**.

86.   Represa, A.; Deloulme, J. C.; Sensenbrenner, M.; Ben-Ari, Y.; Baudier, J., Neurogranin: immunocytochemical localization of a brain-specific protein kinase C substrate. *J Neurosci* **1990,** *10* (12), 3782-92.

87.   Utal, A. K.; Stopka, A. L.; Roy, M.; Coleman, P. D., PEP-19 immunohistochemistry defines the basal ganglia and associated structures in the adult human brain, and is dramatically reduced in Huntington's disease. *Neuroscience* **1998,** *86* (4), 1055-63.

88.   Wotjak, C. T.; Kubota, M.; Liebsch, G.; Montkowski, A.; Holsboer, F.; Neumann, I.; Landgraf, R., Release of vasopressin within the rat paraventricular nucleus in response to

emotional stress: a novel mechanism of regulating adrenocorticotropic hormone secretion? *J Neurosci* **1996,** *16* (23), 7725-32.

89.   Cort, J. H.; Schuck, O.; Stribrna, J.; Skopkova, J.; Jost, K.; Mulder, J. L., Role of the disulfide bridge and the C-terminal tripeptide in the antidiuretic action of vasopressin in man and the rat. *Kidney Int* **1975,** *8* (5), 292-302.

90.   Chauvet, M. T.; Chauvet, J.; Acher, R., The glycopeptide domain of the rat vasopressin precursor. *FEBS Lett* **1983,** *163* (2), 257-60.

91.   Chauvet, M. T.; Chauvet, J.; Acher, R., Guinea pig copeptin. The glycopeptide domain of the vasopressin precursor. *FEBS Lett* **1986,** *197* (1-2), 169-72.

92.   Morgenthaler, N. G.; Struck, J.; Alonso, C.; Bergmann, A., Assay for the measurement of copeptin, a stable peptide derived from the precursor of vasopressin. *Clin Chem* **2006,** *52* (1), 112-9.

93.   Ivell, R.; Schmale, H.; Richter, D., Glycosylation of the arginine vasopressin/neurophysin II common precursor. *Biochem Biophys Res Commun* **1981,** *102* (4), 1230-6.

94.   Acher, R.; Chauvet, J.; Rouille, Y., Dynamic processing of neuropeptides: sequential conformation shaping of neurohypophysial preprohormones during intraneuronal secretory transport. *J Mol Neurosci* **2002,** *18* (3), 223-8.

95.   Sloane, A. J.; Duff, J. L.; Wilson, N. L.; Gandhi, P. S.; Hill, C. J.; Hopwood, F. G.; Smith, P. E.; Thomas, M. L.; Cole, R. A.; Packer, N. H.; Breen, E. J.; Cooley, P. W.; Wallace, D. B.; Williams, K. L.; Gooley, A. A., High throughput peptide mass fingerprinting and protein macroarray analysis using chemical printing strategies. *Mol Cell Proteomics* **2002,** *1* (7), 490-9.

96.   Ericsson, D.; Ekstrom, S.; Nilsson, J.; Bergquist, J.; Marko-Varga, G.; Laurell, T., Downsizing proteolytic digestion and analysis using dispenser-aided sample handling and nanovial matrix-assisted laser/desorption ionization-target arrays. *Proteomics* **2001,** *1* (9), 1072-

97.   Holle, A.; Haase, A.; Kayser, M.; Hohndorf, J., Optimizing UV laser focus profiles for improved MALDI performance. *J Mass Spectrom* **2006,** *41* (6), 705-16.

98.   Chaurand, P.; Sanders, M. E.; Jensen, R. A.; Caprioli, R. M., Proteomics in Diagnostic Pathology: Profiling and Imaging Proteins Directly in Tissue Sections. *Am. J. Pathol.* **2004,** *165* (4), 1057-1068.

99.   Chaurand, P.; Schwartz, S. A.; Reyzer, M. L.; Caprioli, R. M., Imaging mass spectrometry: principles and potentials. *Toxicol Pathol* **2005,** *33* (1), 92-101.

100.   Lemhire, R.; Menguellet, S. A.; Stauber, J.; Marchaudon, V.; Lucot, J. P.; Collinet, P.; Farine, M. O.; Vinatier, D.; Day, R.; Ducoroy, P.; Salzet, M.; Fournier, I., Specific MALDI imaging and profiling for biomarker hunting and validation: Fragment of the 11S proteasome activator

complex, reg alpha fragment, is a new potential ovary cancer biomarker. *Journal of Proteome Research* **2007,** *6*, 4127-4134.

101.	Caprioli, R. M., Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Research* **2005,** *65* (23), 10642-10645.

102.	Fox, C. H.; Johnson, F. B.; Whiting, J.; Roller, P. P., Formaldehyde fixation. *J Histochem Cytochem* **1985,** *33* (8), 845-53.

103.	Ronci, M.; Bonanno, E.; Colantoni, A.; Pieroni, L.; Di Ilio, C.; Spagnoli, L. G.; Federici, G.; Urbani, A., Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations. *Proteomics* **2008,** *8* (18), 3702-14.

104.	Sompuram, S. R.; Vani, K.; Messana, E.; Bogen, S. A., A molecular mechanism of formalin fixation and antigen retrieval. *Am J Clin Pathol* **2004,** *121* (2), 190-9.

105.	Jiang, X.; Feng, S.; Tian, R.; Ye, M.; Zou, H., Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues. *J Proteome Res* **2007,** *6* (3), 1038-47.

106.	Hwang, S. I.; Thumar, J.; Lundgren, D. H.; Rezaul, K.; Mayya, V.; Wu, L.; Eng, J.; Wright, M. E.; Han, D. K., Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene* **2007,** *26* (1), 65-76.

107.	Crockett, D. K.; Lin, Z. S.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S. J., Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab Invest* **2005,** *85* (11), 1405-1415.

108.	Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Furusato, B.; Lucas, D. A.; Ringeisen, B. R.; Sesterhenn, I. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B., Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol Cell Proteomics* **2005,** *4* (11), 1741-1753.

109.	Palmer-Toy, D. E.; Krastins, B.; Sarracino, D. A.; Nadol, J. B.; Merchant, S. N., Efficient method for the proteomic analysis of fixed and embedded tissues. *J Prot Res* **2005,** *4* (6), 2404-2411.

110.	Ahram, M.; Flaig, M. J.; Gillespie, J. W.; Duray, P. H.; Linehan, W. M.; Ornstein, D. K.; Niu, S.; Zhao, Y.; Petricoin, E. F., 3rd; Emmert-Buck, M. R., Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. *Proteomics* **2003,** *3* (4), 413-21.

111.	Shi, S. R.; Liu, C.; Balgley, B. M.; Lee, C.; Taylor, C. R., Protein extraction from formalin-fixed, paraffin-embedded tissue sections: Quality evaluation by mass spectrometry. *J Histochem Cytochem* **2006,** *54* (6), 739-743.

112.   Kononen, J.; Bubendorf, L.; Kallioniemi, A.; Barlund, M.; Schraml, P.; Leighton, S.; Torhorst, J.; Mihatsch, M. J.; Sauter, G.; Kallioniemi, O. P., Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature medicine* **1998,** *4* (7), 844-7.

113.   Battifora, H., The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing. *Lab Invest* **1986,** *55* (2), 244-8.

114.   Battifora, H.; Mehta, P., The checkerboard tissue block. An improved multitissue control block. *Lab Invest* **1990,** *63* (5), 722-4.

115.   Groseclose, M. R.; Massion, P. P.; Chaurand, P.; Caprioli, R. M., High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics* **2008,** *8* (18), 3715-24.

116.   Conrad, D. H.; Goyette, J.; Thomas, P. S., Proteomics as a method for early detection of cancer: a review of proteomics, exhaled breath condensate, and lung cancer screening. *J Gen Intern Med* **2008,** *23 Suppl 1*, 78-84.

117.   Beer, D. G.; Kardia, S. L.; Huang, C. C.; Giordano, T. J.; Levin, A. M.; Misek, D. E.; Lin, L.; Chen, G.; Gharib, T. G.; Thomas, D. G.; Lizyness, M. L.; Kuick, R.; Hayasaka, S.; Taylor, J. M.; Iannettoni, M. D.; Orringer, M. B.; Hanash, S., Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature medicine* **2002,** *8* (8), 816-24.

118.   Bhattacharjee, A.; Richards, W. G.; Staunton, J.; Li, C.; Monti, S.; Vasa, P.; Ladd, C.; Beheshti, J.; Bueno, R.; Gillette, M.; Loda, M.; Weber, G.; Mark, E. J.; Lander, E. S.; Wong, W.; Johnson, B. E.; Golub, T. R.; Sugarbaker, D. J.; Meyerson, M., Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* **2001,** *98* (24), 13790-5.

119.   Garber, M. E.; Troyanskaya, O. G.; Schluens, K.; Petersen, S.; Thaesler, Z.; Pacyna-Gengelbach, M.; van de Rijn, M.; Rosen, G. D.; Perou, C. M.; Whyte, R. I.; Altman, R. B.; Brown, P. O.; Botstein, D.; Petersen, I., Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* **2001,** *98* (24), 13784-9.

120.   Kargi, A.; Gurel, D.; Tuna, B., The diagnostic value of TTF-1, CK 5/6, and p63 immunostaining in classification of lung carcinomas. *Appl Immunohistochem Mol Morphol* **2007,** *15* (4), 415-20.

121.   Khayyata, S.; Yun, S.; Pasha, T.; Jian, B.; McGrath, C.; Yu, G.; Gupta, P.; Baloch, Z., Value of P63 and CK5/6 in distinguishing squamous cell carcinoma from adenocarcinoma in lung fine-needle aspiration specimens. *Diagn Cytopathol* **2009**.

122.   Sandler, A., Bevacizumab in non small cell lung cancer. *Clin Cancer Res* **2007,** *13* (15 Pt 2), s4613-6.

123.   Herbst, R. S.; O'Neill, V. J.; Fehrenbacher, L.; Belani, C. P.; Bonomi, P. D.; Hart, L.; Melnyk, O.; Ramies, D.; Lin, M.; Sandler, A., Phase II study of efficacy and safety of bevacizumab in

combination with chemotherapy or erlotinib compared with chemotherapy alone for treatment of recurrent or refractory non small-cell lung cancer. *J Clin Oncol* **2007,** *25* (30), 4743-50.

124.   Tusher, V. G.; Tibshirani, R.; Chu, G., Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **2001,** *98* (9), 5116-21.

125.   Fraire, A. E.; Roggli, V. L.; Vollmer, R. T.; Greenberg, S. D.; McGavran, M. H.; Spjut, H. J.; Yesner, R., Lung cancer heterogeneity. Prognostic implications. *Cancer* **1987,** *60* (3), 370-5.

126.   Barak, V.; Goike, H.; Panaretakis, K. W.; Einarsson, R., Clinical utility of cytokeratins as tumor markers. *Clin Biochem* **2004,** *37* (7), 529-40.

127.   Moll, R.; Franke, W. W.; Schiller, D. L.; Geiger, B.; Krepler, R., The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **1982,** *31* (1), 11-24.

128.   Ding, S. J.; Li, Y.; Tan, Y. X.; Jiang, M. R.; Tian, B.; Liu, Y. K.; Shao, X. X.; Ye, S. L.; Wu, J. R.; Zeng, R.; Wang, H. Y.; Tang, Z. Y.; Xia, Q. C., From proteomic analysis to clinical significance: overexpression of cytokeratin 19 correlates with hepatocellular carcinoma metastasis. *Mol Cell Proteomics* **2004,** *3* (1), 73-81.

129.   Oshima, R. G.; Baribault, H.; Caulin, C., Oncogenic regulation and function of keratins 8 and 18. *Cancer Metastasis Rev* **1996,** *15* (4), 445-71.

130.   Hendrix, M. J.; Seftor, E. A.; Chu, Y. W.; Trevor, K. T.; Seftor, R. E., Role of intermediate filaments in migration, invasion and metastasis. *Cancer Metastasis Rev* **1996,** *15* (4), 507-25.

131.   Larcher, F.; Bauluz, C.; Diaz-Guerra, M.; Quintanilla, M.; Conti, C. J.; Ballestin, C.; Jorcano, J. L., Aberrant expression of the simple epithelial type II keratin 8 by mouse skin carcinomas but not papillomas. *Mol Carcinog* **1992,** *6* (2), 112-21.

132.   Vaidya, M. M.; Borges, A. M.; Pradhan, S. A.; Bhisey, A. N., Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa. *Eur J Cancer B Oral Oncol* **1996,** *32B* (5), 333-6.

133.   Fukunaga, Y.; Bandoh, S.; Fujita, J.; Yang, Y.; Ueda, Y.; Hojo, S.; Dohmoto, K.; Tojo, Y.; Takahara, J.; Ishida, T., Expression of cytokeratin 8 in lung cancer cell lines and measurement of serum cytokeratin 8 in lung cancer patients. *Lung Cancer* **2002,** *38* (1), 31-8.

134.   Downey, P.; Cummins, R.; Moran, M.; Gulmann, C., If it's not CK5/6 positive, TTF-1 negative it's not a squamous cell carcinoma of lung. *APMIS* **2008,** *116* (6), 526-9.

135.   Camilo, R.; Capelozzi, V. L.; Siqueira, S. A.; Del Carlo Bernardi, F., Expression of p63, keratin 5/6, keratin 7, and surfactant-A in non-small cell lung carcinomas. *Hum Pathol* **2006,** *37* (5), 542-6.

136.   Carella, R.; Deleonardi, G.; D'Errico, A.; Salerno, A.; Egarter-Vigl, E.; Seebacher, C.; Donazzan, G.; Grigioni, W. F., Immunohistochemical panels for differentiating epithelial malignant mesothelioma from lung adenocarcinoma: a study with logistic regression analysis. *Am J Surg Pathol* **2001,** *25* (1), 43-50.

137.   Ordonez, N. G., The diagnostic utility of immunohistochemistry in distinguishing between epithelioid mesotheliomas and squamous carcinomas of the lung: a comparative study. *Mod Pathol* **2006,** *19* (3), 417-28.

138.   Miettinen, M.; Sarlomo-Rikala, M., Expression of calretinin, thrombomodulin, keratin 5, and mesothelin in lung carcinomas of different types: an immunohistochemical analysis of 596 tumors in comparison with epithelioid mesotheliomas of the pleura. *Am J Surg Pathol* **2003,** *27* (2), 150-8.

139.   Johansson, L., Histopathologic classification of lung cancer: Relevance of cytokeratin and TTF-1 immunophenotyping. *Ann Diagn Pathol* **2004,** *8* (5), 259-67.

140.   Park, S. Y.; Kim, B. H.; Kim, J. H.; Lee, S.; Kang, G. H., Panels of immunohistochemical markers help determine primary sites of metastatic adenocarcinoma. *Arch Pathol Lab Med* **2007,** *131* (10), 1561-7.

141.   Bodenmuller, H.; Ofenloch-Hahnle, B.; Lane, E. B.; Dessauer, A.; Bottger, V.; Donie, F., Lung cancer-associated keratin 19 fragments: development and biochemical characterisation of the new serum assay Enzymun-Test CYFRA 21-1. *Int J Biol Markers* **1994,** *9* (2), 75-81.

142.   Brabon, A. C.; Williams, J. F.; Cardiff, R. D., A monoclonal antibody to a human breast tumor protein released in response to estrogen. *Cancer Res* **1984,** *44* (6), 2704-10.

143.   Ku, N. O.; Liao, J.; Omary, M. B., Apoptosis generates stable fragments of human type I keratins. *J Biol Chem* **1997,** *272* (52), 33197-203.

144.   Nisman, B.; Lafair, J.; Heching, N.; Lyass, O.; Baras, M.; Peretz, T.; Barak, V., Evaluation of tissue polypeptide specific antigen, CYFRA 21-1, and carcinoembryonic antigen in nonsmall cell lung carcinoma: does the combined use of cytokeratin markers give any additional information? *Cancer* **1998,** *82* (10), 1850-9.

145.   Pujol, J. L.; Grenier, J.; Daures, J. P.; Daver, A.; Pujol, H.; Michel, F. B., Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. *Cancer Res* **1993,** *53* (1), 61-6.

146.   Buccheri, G.; Ferrigno, D., Usefulness of tissue polypeptide antigen in staging, monitoring, and prognosis of lung cancer. *Chest* **1988,** *93* (3), 565-70.

147.   Sun, S. S.; Hsieh, J. F.; Tsai, S. C.; Ho, Y. J.; Lee, J. K.; Kao, C. H., Cytokeratin fragment 19 and squamous cell carcinoma antigen for early prediction of recurrence of squamous cell lung carcinoma. *Am J Clin Oncol* **2000,** *23* (3), 241-3.

148.   Vollmer, R. T.; Govindan, R.; Graziano, S. L.; Gamble, G.; Garst, J.; Kelley, M. J.; Christenson, R. H., Serum CYFRA 21-1 in advanced stage non-small cell lung cancer: an early measure of response. *Clin Cancer Res* **2003,** *9* (5), 1728-33.

149.   Holdenrieder, S.; Stieber, P.; von Pawel, J.; Raith, H.; Nagel, D.; Feldmann, K.; Seidel, D., Circulating nucleosomes predict the response to chemotherapy in patients with advanced non-small cell lung cancer. *Clin Cancer Res* **2004,** *10* (18 Pt 1), 5981-7.

150.   Pujol, J. L.; Molinier, O.; Ebert, W.; Daures, J. P.; Barlesi, F.; Buccheri, G.; Paesmans, M.; Quoix, E.; Moro-Sibilot, D.; Szturmowicz, M.; Brechot, J. M.; Muley, T.; Grenier, J., CYFRA 21-1 is a prognostic determinant in non-small-cell lung cancer: results of a meta-analysis in 2063 patients. *Br J Cancer* **2004,** *90* (11), 2097-105.

151.   Stieber, P.; Dienemann, H.; Hasholzner, U.; Muller, C.; Poley, S.; Hofmann, K.; Fateh-Moghadam, A., Comparison of cytokeratin fragment 19 (CYFRA 21-1), tissue polypeptide antigen (TPA) and tissue polypeptide specific antigen (TPS) as tumour markers in lung cancer. *Eur J Clin Chem Clin Biochem* **1993,** *31* (10), 689-94.

152.   Rastel, D.; Ramaioli, A.; Cornillie, F.; Thirion, B., CYFRA 21-1, a sensitive and specific new tumour marker for squamous cell lung cancer. Report of the first European multicentre evaluation. CYFRA 21-1 Multicentre Study Group. *Eur J Cancer* **1994,** *30A* (5), 601-6.

153.   Lagarde, A.; Forgach, T.; Nagy, D.; Nagy, K.; Vasas, S.; Janoki, G. A., Diagnostic sensitivity of three tumour markers in non-small cell lung cancer: a pilot study. *Nucl Med Rev Cent East Eur* **2000,** *3* (2), 139-42.

154.   Miedouge, M.; Devys, A.; Simon, M.; Rouzaud, P.; Salama, G.; Reyre, J.; Pujazon, M.; Carles, P.; Serre, G., High levels of cytokeratin 19 fragments but no evidence of cytokeratins 1, 2, 10/11, 14 or filaggrin in the serum of squamous cell lung carcinoma patients. *Tumour Biol* **2001,** *22* (1), 19-26.

155.   Pujol, J. L.; Grenier, J.; Parrat, E.; Lehmann, M.; Lafontaine, T.; Quantin, X.; Michel, F. B., Cytokeratins as serum markers in lung cancer: a comparison of CYFRA 21-1 and TPS. *Am J Respir Crit Care Med* **1996,** *154* (3 Pt 1), 725-33.

156.   Rapellino, M.; Piantino, P.; Pecchio, F.; Ruffini, E.; Cavallo, A.; Scappaticci, E.; Baldi, S.; Ciocia, E.; Pivetti, S., Disappearance curves of tumor markers after radical surgery. *Int J Biol Markers* **1994,** *9* (1), 33-7.

157.   Yen, T. C.; Lin, W. Y.; Kao, C. H.; Cheng, K. Y.; Wang, S. J., A study of a new tumour marker, CYFRA 21-1, in squamous cell carcinoma of the head and neck, and comparison with squamous cell carcinoma antigen. *Clin Otolaryngol Allied Sci* **1998,** *23* (1), 82-6.

158.   Ebert, W.; Muley, T.; Drings, P., Does the assessment of serum markers in patients with lung cancer aid in the clinical decision making process? *Anticancer Res* **1996,** *16* (4B), 2161-8.

159.   Marchi, N.; Mazzone, P.; Fazio, V.; Mekhail, T.; Masaryk, T.; Janigro, D., ProApolipoprotein A1: a serum marker of brain metastases in lung cancer patients. *Cancer* **2008,** *112* (6), 1313-24.

160.   Takenaka, K.; Ogawa, E.; Oyanagi, H.; Wada, H.; Tanaka, F., Carbonyl reductase expression and its clinical significance in non-small-cell lung cancer. *Cancer Epidemiol Biomarkers Prev* **2005,** *14* (8), 1972-5.

161.   Sung, H. J.; Cho, J. Y., Biomarkers for the lung cancer diagnosis and their advances in proteomics. *BMB Rep* **2008,** *41* (9), 615-25.

162.   Yao, H.; Zhang, Z.; Xiao, Z.; Chen, Y.; Li, C.; Zhang, P.; Li, M.; Liu, Y.; Guan, Y.; Yu, Y.; Chen, Z., Identification of metastasis associated proteins in human lung squamous carcinoma using two-dimensional difference gel electrophoresis and laser capture microdissection. *Lung Cancer* **2008**.

163.   Huang, Y.; Jin, Y.; Yan, C. H.; Yu, Y.; Bai, J.; Chen, F.; Zhao, Y. Z.; Fu, S. B., Involvement of Annexin A2 in p53 induced apoptosis in lung cancer. *Mol Cell Biochem* **2008,** *309* (1-2), 117-23.

164.   Rakotomalala, R., TANAGRA: a free software for research and academic purposes. *Proceedings of EGC'2005, RNTI-E-3* **2005,** *2*, 697-702.

165.   Seo, J.; Gordish-Dressman, H.; Hoffman, E. P., An interactive power analysis tool for microarray hypothesis testing and generation. *Bioinformatics* **2006,** *22* (7), 808-14.