A MOLECULAR ASSESSMENT OF TUMOR MARGINS BY MALDI MASS SPECTROMETRY IN RENAL CARCINOMA

By

Stacey Renee Oppenheimer

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

May, 2007

Nashville, Tennessee

Approved:

Professor Richard M. Caprioli

Professor Carlos L. Arteaga

Professor Robert J. Matusik

Professor Michael P. Stone

Professor Jennifer A. Pietenpol

This work is dedicated to my parents, John and Toni, whose love and support have helped me persevere and accomplish my ambitions.

This work is also dedicated in honor of Maud, my grandmother, and in loving memory of Henry, my grandfather.

ACKNOWLEDGEMENTS

Always bear in mind that your own resolution to succeed is more important than any one thing.

—Abraham Lincoln

Having reached the end of a long road that has been challenging both mentally and emotionally, I can now reflect on the factors that have truly influenced my journey. I have been blessed to be surrounded by wonderful people, to whom I owe many thanks.

First I would like to thank my research advisor, Richard Caprioli, for believing in me, presenting me with the challenge, and allowing me to meet that challenge. Thank you for your support and guidance as a mentor. Thanks also to remaining members of my Ph.D. committee, Robert Matusik, Michael Stone, Jennifer Pietenpol and Carlos Arteaga, for their guidance and support.

Next, I would like to thank all of the people who have had an influence on the progress of this project. First and foremost, thanks to Deming Mi and Melinda Sanders. I am grateful to Deming for providing the statistical expertise that made this project a success. I am thankful to Melinda Sanders, the pathologist, for taking on my project and providing helpful suggestions. Additionally, I would like to thank Julia Grigorieva and Heinrich Roder at Biodesix, Inc. for their statistical analysis. Thanks to Scott Sobecki and Shannon Cornett for custom programs. Also, thanks to Kristin Burnum for her guidance with protein identification, Lisa Zimmerman and Julie Coleman in the Proteomics Core facility for access to the IPG-separated peptide solutions used for protein identification and Malin Andersson for her assistance with immunostaining. Lastly, thank you to all members of the Mass Spectrometry Research Center who have helped me over the years, either through assisting with an experiment or providing insightful ideas.

I would also like to acknowledge the sources of tissue samples and funding. Tissue samples were collected from the Human Tissue Acquisition and Pathology Core laboratory at the Vanderbilt University Medical Center, the Cooperative Human Tissue Network, and the Fox Chase Cancer Center Tumor Repository. Funding was provided by NIH/NIGMS 5R01 GM 58008, NIH/NCI 5R33 CA86243, and DOD W81XWH-05-01-0179, and I would like to acknowledge support from the Training Program in Environmental Toxicology T32 ES07028-29.

On a personal note, I would like to thank all current and former members of the Mass Spectrometry Research Center who have supported me over the years. You were not just colleagues but also friends. Sheerin Shahidi, Kristen Herring, and Kristin Burnum, you are the sisters I never had. Maureen Casey, thank you for being a surrogate mother to me and the other students in Dr. Caprioli's lab.

Unfortunately, teachers do not receive enough credit for the influences they make on a student's life. It is my honor to acknowledge the three individuals who had the most influence on my decision to pursue a higher level of chemistry education. Thanks to Charles Parker, my high school chemistry teacher, for making chemistry challenging, but enjoyable enough to pursue further. Tracie Williams was my analytical chemistry professor in college. Although she was only there for six months, she left a lasting impression, instilling in me her passion for mass spectrometry. John Shibata, a physical chemist, was probably the most patient and encouraging professor I ever had. John and Tracie, thank you for your influence and continued friendship.

Lastly, thank you to my family, including John, Toni, my grandmothers, Regina, Barry, Darren, Beth, and Caroline. You are the most loving and supportive family, and I would not have made it this far without you. To my parents, thank you for your endless faith and support. Thanks for raising me to be the person I am today. I love you and may God bless you always.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AJCC	American Joint Committee on Cancer
ARM	Acoustic reagent multispotter
AUC	Area under the curve
ccRCC	Clear cell renal cell carcinoma
CHIP	Chemical inkjet printer
DIGE	Difference in-gel electrophoresis
FDR	False discovery rate
FTICR	Fourier transform ion cyclotron resonance
IEF	Isoelectric focusing
IHC	Immunohistochemistry
ITO	Indium tin oxide
IUAC	International Union Against Cancer
LOOCV	Leave-one-out cross validation
LOWESS	Locally-weighted-regression scatter-plot smoothing
LC	Liquid chromatography
m/z	Mass-to-charge
MALDI	Matrix-assisted laser desorption ionization
MDS	Multidimensional scaling
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS	Mass spectrometry
NSS	Nephron sparing surgery
OCT	Optimal cutting temperature
PAGE	Polyacrylamide gel electrophoresis
PMF	Peptide mass fingerprint

RCC	Renal cell carcinoma
rf	Radiofrequency
RT-PCR	Real-time polymerase chain reaction
SA	Sinapinic acid
SAM	Significance analysis of microarrays
SIMS	Secondary ion mass spectrometry
SVM	Support vector machine
TFA	Trifluoroacetic acid
TIC	Total ion-current
TOF	Time-of-flight
TNM	Tumor, nodes, metastases
VHL	von Hippel-Lindau

CHAPTER I

BACKGROUND AND OBJECTIVES

The wealth of information offered to the life sciences by the success of the Human Genome Project[4, 5] has lead to the revolutionary field of proteomics, the analysis of all proteins in a living system, which includes their post-translational modifications, spliced variants, covalent and non-covalent associations, spatial and temporal distributions within cells and how these are affected by changes in their environment[6]. The thousands or more proteins encoded by the genome are involved in numerous activities within tissue, and their expression levels and molecular forms are a consequence of genomic factors, post-translational modifications, regulatory processes, environmental factors and temporal processes[7, 8]. The net sum of these gives rise to a proteome expression level and distribution that reflects the integrated metabolic state of the cells in that tissue at any given time, such as in diseased or normal tissue.

Advanced technologies, including gel electrophoresis and mass spectrometry, are now used to facilitate a better understanding of human disease through proteomic analyses. Numerous collaborative efforts are underway, such as the Human Proteome Organization[9, 10], to discover arrays of proteins that may serve as disease markers or therapeutic targets. Other organizations, such as the Human Proteome Atlas[11] and the NCI Clinical Proteomics Technologies Initiative, specifically target cancer biomarkers. Despite these endeavors, little effort has been focused on the proteomics of cellular environments around tumor margins. Matrix-assisted laser desorption/ionization time-of-

flight mass spectrometry (MALDI-TOF MS) is one of the few proteomic technologies that provides the precise spatial information needed to examine the proteomics of tumor progression and invasion. The application of MALDI-TOF MS to the proteomic analysis of tumor margins will be discussed in this chapter.

Clinical Proteomics

There are several subsets of proteomics research. Many efforts focus on the large-scale identification of proteins and their posttranslational modifications within a cell, tissue, or other biological sample (urine, serum, etc.) followed by characterization. The subset of structural proteomics involves determining the structure of these individually characterized proteins, while functional proteomics focuses on identifying protein functions and their biological interactions (pathways)[12]. Though many research efforts remain focused on hypothesis-driven projects involving single or classes of proteins and their pathways, the field now includes discovery driven research. This field uses high-throughput approaches such as 2D gel electrophoresis and mass spectrometry to compare the proteome expression levels and distributions of biological systems at any given point, such as different disease states.

These high-throughput approaches are used in the recently emerged field of clinical proteomics, which deals with the hypothesis that proteins can provide diagnostic and/or therapeutic solutions to cancer and other diseases, focusing on the comparison of the differential display of protein levels in different biological states (e.g. normal versus cancer). In differential display proteomics, one would hope to find unique markers (biomarkers), perhaps resulting from posttranslational modifications of proteins in the

cancerous cells or loss-of-function proteins, which are not present in one state in relation to the other. Complete loss of function, however, is not always observed, thus most studies now aim to discover proteins that are significantly up-regulated or down-regulated in the test system as compared to the control. It is also unlikely that a single protein biomarker will provide the sensitivity and specificity required for disease detection and prognosis; thus the focus is shifting from methods that analyze one marker at a time to high-throughput protein pattern-matching approaches, such as 2D gel electrophoresis and mass spectrometry, that facilitate the simultaneous measurement of a range of possible disease markers directly related to disease and cancer processes[7, 12, 13].

Update on Clinical Proteomic Technologies

Gel Electrophoresis Approaches

Polyacrylamide gel electrophoresis

Conventionally, 2D polyacrylamide gel electrophoresis (2D PAGE) has been the protein separation technology most associated with proteomics and biomarker discovery. 2D PAGE has been widely used over the last 40 years to resolve several thousand proteins from a single sample. In the first dimension of this 2D technology, proteins undergo isoelectric focusing (IEF), which separates biological species according to their net charge, or isoelectric point. During the second dimension, which is orthogonal to the first, proteins are separated by molecular weight. Gels are then stained with a visible or fluorescent dye, allowing for visualization of protein spots on the gel. Protein spots of

interest are then excised from the gel, digested, analyzed by mass spectrometry and identified by searching theoretical sequence databases. This technology has facilitated the identification of many proteins in tissues or subcellular fractions. For the field of clinical proteomics, 2D PAGE has been used to compare the relative abundances of proteins in related samples, such as differentiation between prostate cancer and benign prostate hyperplasia[14], analysis of cerebrospinal fluid to find commonalities between Schizophrenia and Alzheimer's disease[15], determining markers of chronic lymphocytic leukaemia[16] and of metastatic head and neck squamous cell carcinoma[17].

Despite its contributions, this technology has some major limitations in its ability to contribute to clinical proteomics. One pitfall is that it only produces qualitative data, thus providing no quantitative information relating to the extent of change in particular protein between disease states. A second drawback of the 2D PAGE approach is its considerable gel-to-gel variability. Lastly, normalization of gels is difficult to achieve because the most commonly used stain, coomassie blue, is protein dependent. The sum of the factors adds too much variability to the system, rendering it inadequate for differential analyses desired for clinical proteomics[18].

Two-dimensional difference in-gel electrophoresis

Circumventing the drawbacks of conventional 2D PAGE technology, 2D difference in-gel electrophoresis (DIGE)[19] offers high reproducibility and an increased dynamic range. This technology uses pre-electrophoretic labeling of samples with one of three fluorescent CyDyes (Cy2, 3, and 5). Two CyDye labels are available: N-hydroxy succinimidyl ester reagents that label the ε-amine groups of lysine residues and free N-

termini, and maleimide reagents for labeling cysteine sulfhydryls. In a typical DIGE experiment, the protein mixtures to be compared are labeled with either Cy3 (e.g. cancer) or Cy5 (e.g. normal). The Cy2 is used on a pooled sample comprising equal amounts of each protein mixture[20]. This internal standard is used to correct for inter- and intra-gel variability. From the image information, spot volumes from each dye are calculated and compared. The presence of the same pooled standard on each gel allows for various statistical analyses despite having samples separated on different DIGE gels.

This technology has been successfully applied to various clinical proteomics studies, including human breast cancer[21], human esophageal cancer[22], and human colon cancer[23]. Although 2D gel electrophoresis has the highest resolving power of any multidimensional proteomic separation methods at up to 10,000 proteins, its limited loading capacity hampers its ability to identify low-abundant proteins. There are also difficulties in analyzing proteins with extreme isoelectric points, molecular weights (optimum 20,000 to 100,000 Da) and hydophobicities[24]. Although 2D gel electrophoresis separation technologies followed by mass spectrometry identification has been the most commonly used method for protein separation and protein identification, this method of proteome analysis remains challenging due to the individual extraction, digestion and analysis of each gel spot, which is a tedious and time-consuming process.

Mass Spectrometry Approaches

Multidimensional liquid chromatography-tandem mass spectrometry

Many efforts have been underway to utilize multidimensional liquid chromatography (LC) approaches coupled to mass spectrometry to separate and characterize proteomes. In contrast to the tedious protein identification steps required post 2D gel electrophoresis techniques, this approach has the ability to identify proteins within a single, automated step. In brief, samples of interest (tissues, cells, subcellular fractions, or fluids) are homogenized and then, in most cases, digested into peptides. Peptides are then separated by two orthogonal chromatography steps followed by tandem mass spectrometry, which involves further fragmenting the peptides in order to obtain more database sequence coverage. This method can be done online or off-line. The on-line approach minimizes sample handling, limiting sample loss, because peptides elute from the LC directly into the mass spectrometer. This approach is limited to lower loading capacities. The off-line approach eliminates these major problems, but results in some sample loss. The method used will ultimately depend on the sample to be analyzed and the research objectives[24-28].

While the second dimension separation is conventionally reversed-phase LC, several options are available for the first dimension separation step. Conventional approaches utilize liquid-phase ion-exchange chromatography as the first dimension, which allows for collecting fractions of peptides by varying isoelectric points in an off-line approach. In the second separation, each of the fractions from the first dimension is subjected to reversed phase LC. This step is usually done on-line with a mass

spectrometer to obtain MS/MS information as peptides elute from the column. Other first dimension separations that have been combined with reverse-phase LC include size exclusion chromatography[28], affinity chromatography (e.g. immobilized metal-affinity chromatography)[29], and more recently immobilized pH gradient gel (IPG) isoelectric focusing (IEF). The latter approach offers increased confidence for protein identification by significantly reducing the rate of false positives. More detailed information on IEF peptide separation with IPG gels can be found elsewhere[30-33].

A differential proteomics approach to the multidimensional LC-MS/MS technology is achieved by isotopic labeling. Labeling approaches include reagents, such as ICAT[34] and ITRAQ[35], as well as metabolic labeling and enzymatic methods that include SILAC[36] and trypsin-catalyzed ^{16}O-to-^{18}O exhange[37]. These approaches label specific residues, such as cysteines, lysines, amine- and carboxy-termini, of two different protein mixtures with either a light (e.g. ^{12}C) or heavy (e.g. ^{13}C) isotope tag. Depending on the label, the two mixtures are labeled then digested or digested first and then labeled. The two differentially labeled samples are mixed together and fractionated using multidimensional LC with subsequent relative quantitation of protein or peptide levels and identification by mass spectrometry. Clinical proteomics-based research involving these technologies include markers for persistent Mycobacterium tuberculosis[38], cisplatin resistance in ovarian cancer cells[39], TGF-beta induction of epithelial-mesenchymal transition in lung cancer cells[40], endometrial cancer[41] and metastatic prostate cancer[42]. Although, theoretically, there is an unlimited dynamic range with this technology, most proteins identified are high molecular weight proteins, as their digested peptide fragments tend to saturate the protein mixture analyzed; thus, few low molecular weight (<20,000)

proteins are detected. Many reagents needed for labeling are expensive and the labeling processes can be laborious and time consuming, though this is not always the case.

Accurate mass tag

An alternative, label-free approach has been developed, which uses relative peptide peak intensities, after a normalization procedure, from peptides that have been analyzed by high resolution mass spectrometry with high mass accuracy. The differential quantitative analysis of samples using this technology involves three steps. First, a complex protein mixture from the system of interest is enzymatically digested and analyzed by 2D LC-MS/MS using an FTICR mass spectrometer. The resulting spectra are searched against protein databases, which allows for the creation of a customized protein database with accurate mass and normalized elution time lists of the system being studied. Next, enzymatic digests from the control and experimental state of the system (e.g. normal versus cancerous) are analyzed by LC-FTICR. Peptide identities are determined by searching against the customized database. Finally, abundance ratios are determined using peptides that are consistently identified in replicate experiments. Abundance ratios of peptides and their corresponding proteins are calculated to compare the two proteomes under comparison (e.g. normal versus cancer)[43]. Although this is a potentially invaluable approach to clinical proteomics, it is not practical for most labs, as FT instruments are extremely expensive.

Mass Spectrometry Approaches without Prior Separation

The coupling of 2D gel electrophoresis and LC to mass spectrometry has played a major role in examining differentially expressed proteins in disease, cancer and toxicity studies, but since then, efforts have also focused on employing high-throughput MS technologies to directly analyze human samples (e.g. tissues, sera, urine) for differential proteome analysis of complex samples without prior separation. MALDI-TOF MS is high-throughput and robust. In contrast to the previously mentioned technologies, this approach also requires less tedious sample preparation and requires no prior separation steps. Utilizing MALDI-TOF MS, direct tissue profiling/imaging allows for *in-situ* molecular analysis to investigate the proteome while retaining spatial information, a characteristic lacking in classical proteomic tools.

MALDI-TOF

- Sample Preparation

In preparation for MALDI, the protein mixture of interest is mixed in about a 1:5000 ratio with a small, acidic and energy absorbing organic molecule, called the matrix compound. Typical matrices used are 3,5-dimethoxy-4-hydroxycinnamic acid for proteins (sinapinic acid, SA), α-cyano-4-hydroxycinnamic acid for peptides, and 2,5-dihydroxybenzoic acid for smaller organic molecules such as drugs. The matrix used depends on the chemistry of the analyte such as size, hydrophobicity, and salt content. A drop (~100-500 nL) of the analyte/matrix mixture is then placed onto a conductive MALDI target plate, where the matrix and analyte will then co-crystallize as the solvents evaporate.

- MALDI

After sample preparation, the MALDI target plate is then placed into the mass spectrometer and irradiated with a series of brief laser pulses. The type and wavelength of the laser employed is important because molecular desorption is dependent on the wavelength used. Although most commercial instruments use ultraviolet lasers (commonly, nitrogen or frequency-tripled Nd:YAG), some infrared lasers have been utilized. When the laser strikes the matrix crystals, the matrix crystals absorb the energy to photoionize. The plume of matrix and analyte molecules enter into the gas phase, where ions are formed through gas phase proton-transfer reactions, creating mostly singly charged analyte ions[44-46]. A depiction of MALDI is summarized in Figure 1.

- Time-of-flight mass analyzer

The process of MALDI creates packets of large molecular weight ions in frequent, nanosecond laser pulses, requiring a mass analyzer, such as TOF analyzer[47], illustrated in Figure 2. After ions are created in the MALDI ion source, the ions are accelerated into a field-free drift tube of fixed dimension (L) flanked on one end by a MALDI ion source and the other end by an ion detector[48, 49]. The ion source and drift-tube are operated under high vacuum (10^{-7} to 10^{-8} Torr). Upon entry into the TOF mass analyzer, ions fly through a field-free drift tube prior to impacting the detector.

The total flight time (T) of an ion traveling through the system is a combination of the flight time spent in the ion source (t_s) and the flight time spent in the drift tube (t_L) before hitting the detector.

$$T = t_s + t_L$$

The time ions spend in the source (t_s) is a combination of the mass (m) of the ion and its acceleration. Assuming the desorption process includes negligible initial velocity, t_s can be expressed as:

$$t_s = 2d\sqrt{(m/2qU)}$$

where d and U correspond to the fixed dimension of the ion source and the fixed positive or negative potential bias on the sample plate, respectively. When leaving the source, an ion with mass (m) and charge (q) will have kinetic energy (KE)

$$KE = mv^2/2 = qU$$

In the field-free drift tube of fixed dimension (L), where the velocity (v) of each ion is constant, the time-of-flight (t_L) can be defined as:

$$t_L = L/v = L\sqrt{(m/2qU)}$$

When combining the time spent in the source and the time spent in the drift tube, the total time-of-flight is expressed as:

$$T = (2d + L)\sqrt{(m/2qU)}$$

In summary the total flight time is proportional to the squared-root of the mass-to-charge ratio:

$$T \propto \sqrt{(m/q)}$$

Thus, as the equation suggests, larger molecules will reach the detector slower than smaller molecules. Once the ions hit the detector, spectra are recorded as the mass-to-charge (m/q or m/z) of the ions and their relative intensities[50-52].

As mentioned, the goal of clinical proteomics is to find biomarkers indicative of disease states. All of the aforementioned clinical proteomic technologies have a common disadvantage: regardless of the analyte source, such as primary tumors, all samples must be homogenized. As discussed below, these issues have been overcome by the ability to

analyze biomolecules directly from tissue specimens, without the need for homogenization.

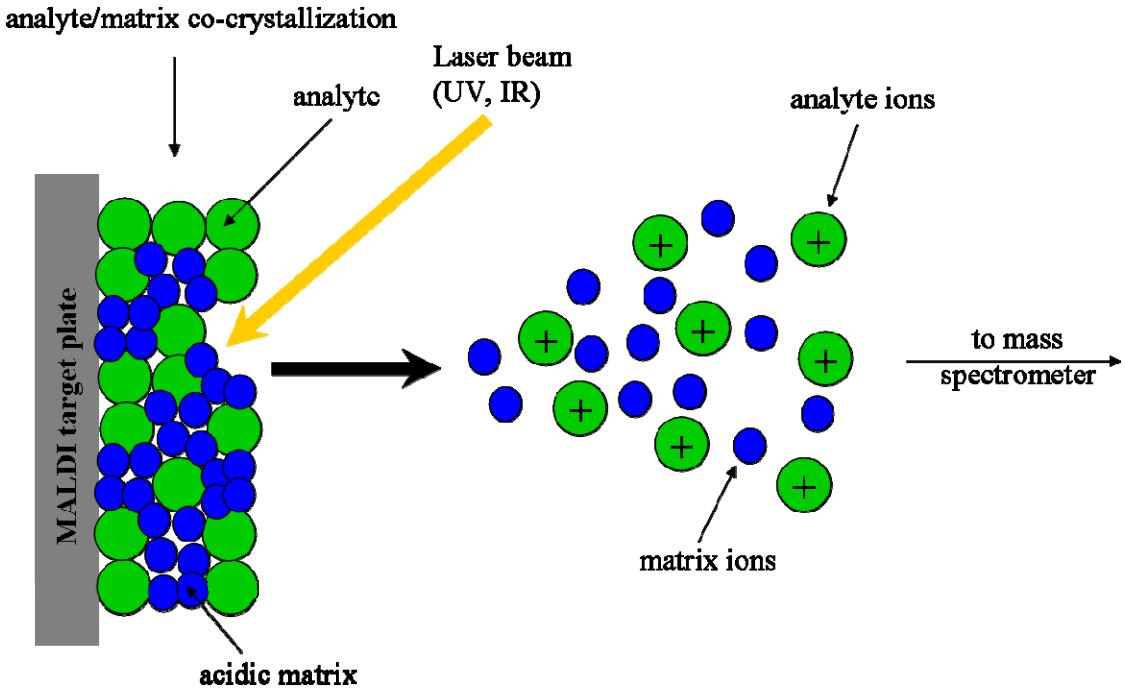


Figure 1: Sample/matrix co-crystallization and ionization. Matrix/analyte co-crystals on the MALDI plate surface are irradiated with a laser pulse. Desorption of the matrix and analyte molecules occurs followed by ionization of the analyte molecules through proton-transfer reactions with the matrix molecules.

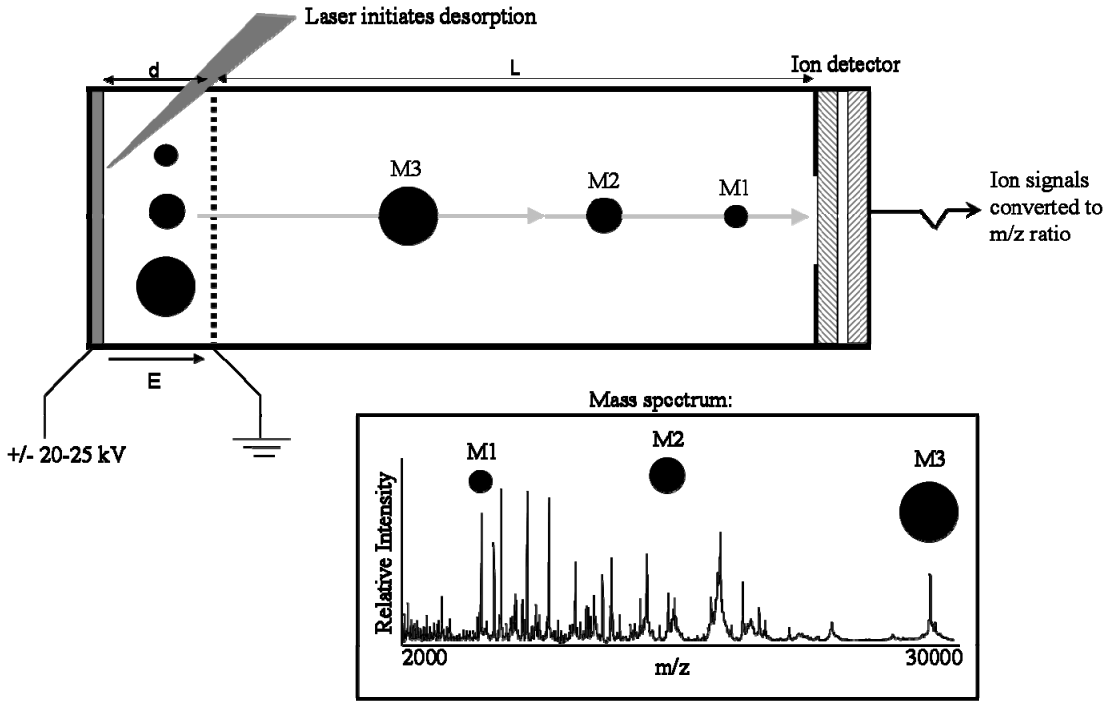


Figure 2: Schematic of a TOF analyzer. The instrument has two parts: 1.) an ion source of fixed length (d) where the desorption/ionization events occur and 2.) a field-free drift tube of fixed length (L). From the source, ions are accelerated through the drift tube and hit the detector. Based on flight time, the mass-to-charge value for each ion is calculated, and the resulting mass spectrum is generated.

Direct Tissue Analysis by Mass Spectrometry for Clinical Proteomics

Direct tissue analysis by MALDI

The ability for MALDI-TOF technology to analyze complex biological samples such as cells and biological fluids has led to its use for direct tissue analysis. This approach eliminates the need for tedious and often time consuming sample preparation steps prior to analysis, offering the ability to analyze proteins and peptides while maintaining their spatial orientation within the tissue. Direct tissue analysis is referred to in two ways: profiling and imaging. As seen in Figure 3, profiling refers to sampling the area of tissue in discrete spots, whereas, imaging is a high resolution profile, in which the laser rasters at a defined spatial resolution across the entire surface of the tissue covered with matrix. The actual ion image is created using custom software that allows visualization of the spatial distribution and relative intensity of a particular ion across the tissue surface.

Direct tissue analysis by MALDI was introduced by Caprioli *et al.*[53] and is beginning to be used by other laboratories[54-58]. This technology has been used to detect drugs and their metabolites[59, 60] as well as intact proteins and peptides[53, 59, 61-66] directly from tissue. Profiling and imaging mass spectrometry have contributed greatly to clinical proteomics. Examples include: Schwartz *et al.* identified prognostic markers for patients with brain tumors[67] as well as potential diagnostic markers that differentiate between different glioma grades[65]; Xu and colleagues identified biomarkers indicative of glomerulosclerosis from laser capture microdissected cells[68]; protein expression changes detected by this technology have been linked to the prediction of tumor response to

molecular therapies[59]; protein profiles generated on an experimental animal model showed its potential in understanding Parkinson's disease progression[69]; Yangisawa and colleagues[70, 71] obtained profiles that were able to classify and predict histological groups as well as nodal involvement and survival in resected non-small-cell lung cancer; the effect of TGF-beta in mammary tumor development was also examined with this technology[72]. More recently, direct tissue analysis has shown potential to assess the proteomic characteristics of microenvironments around tumor margins that may go undetected by histological analyses [Figure 4][3].

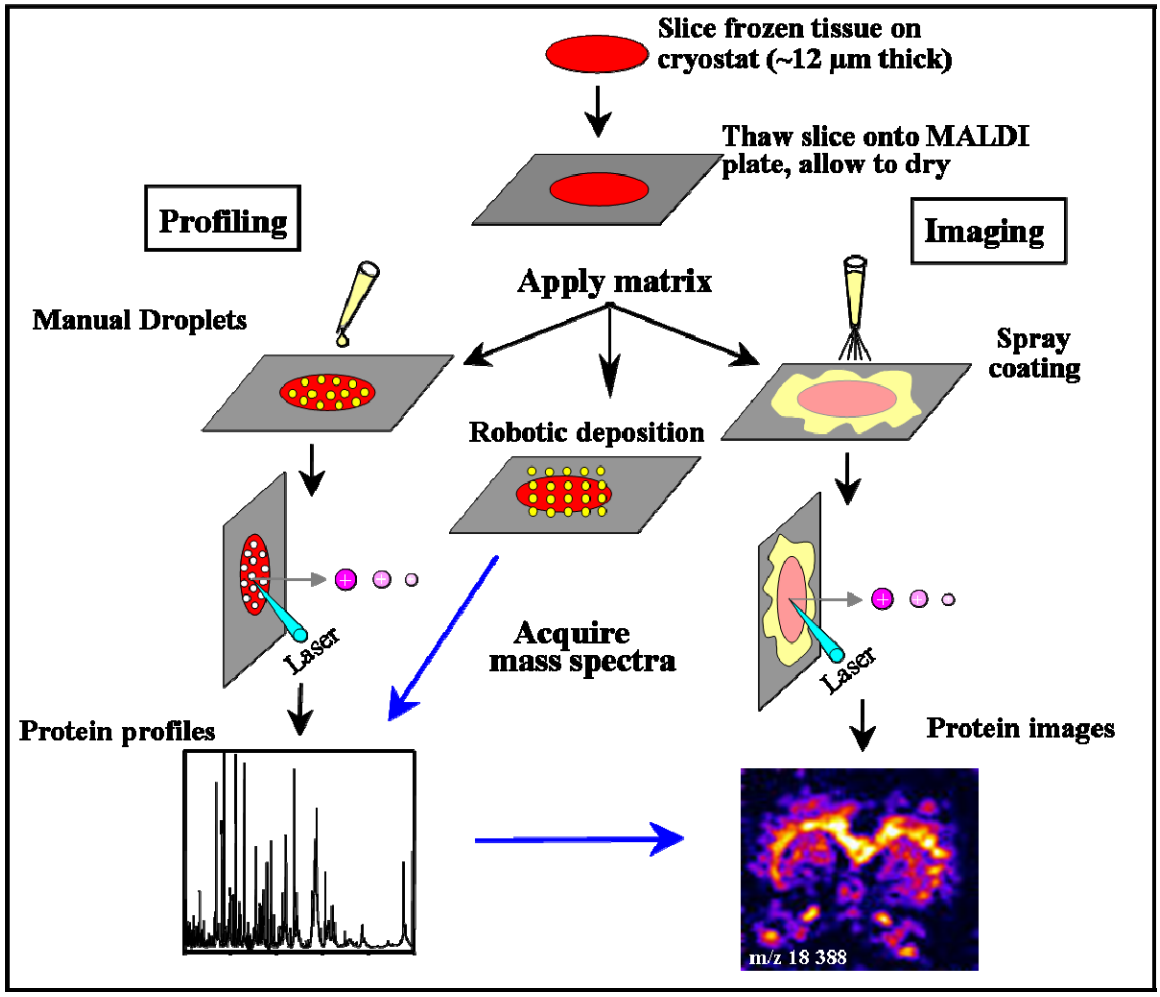


Figure 3: Schematic involved in direct tissue profiling and imaging by MALDI MS. Adapted, with permission, from Caprioli et al.[1]. Copyright 2004 American Chemical Society

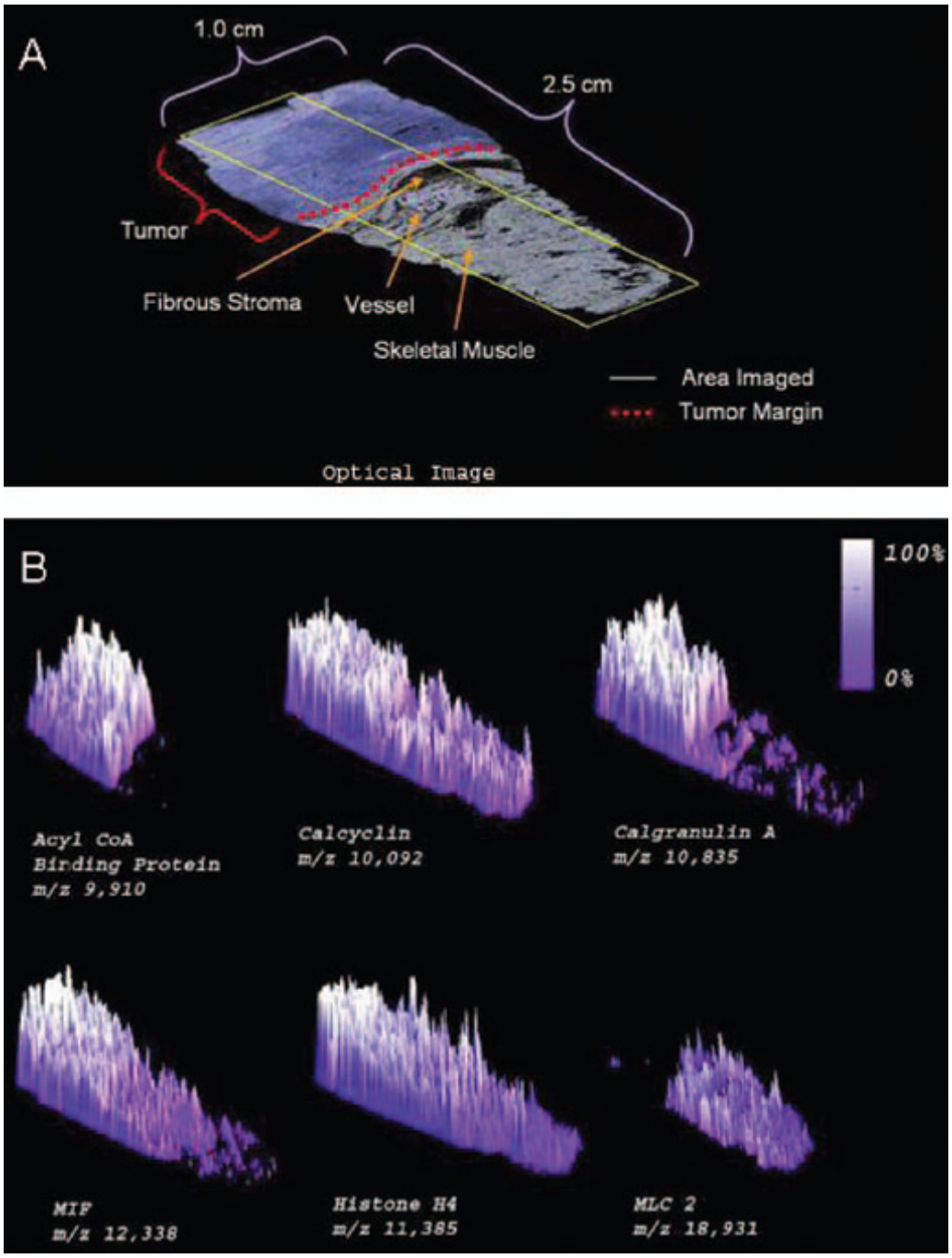


Figure 4: Preliminary analysis of tumor margins by profiling/imaging MALDI MS. This figure illustrates the potential ability of MALDI MS to characterize protein localization in the tumor microenvironment. (A) Optical image of H&E stained image marked by a pathologist. (B) Signals consistent with previously identified proteins are shown. Three-dimensional surface contour representations, created from ion density maps, are based on pixel intensity of the signal of interest. As shown, acyl-CoA binding protein is present only in tumor, whereas MLC 2 is present only in normal. Other proteins demonstrate a gradient-like distribution from tumor into the histologically-defined normal tissue. (Reprinted, with permission, from Caldwell et al.[3])

Direct tissue analysis by SIMS

A second approach to direct tissue analysis is time-of-flight secondary ion mass spectrometry (TOF-SIMS), which is similar to MALDI, but instead of using a pulsed laser, SIMS uses a pulsed primary ion beam (e.g. gallium or indium) to desorb/ionize the sample surface. The resulting "secondary" ions are then accelerated into the TOF analyzer. Primary ion sources are typically operated in the energy range of several hundred eV to several keV. The yield of secondary ions increases with increasing energy, but high energies result in more sample damage. The use of higher mass primary ion sources also enhances the yield of total secondary ions as well as having the capability to increase the secondary ions created in higher mass ranges of the sample[73, 74].

SIMS technology has been around since the 1970s but was primarily used for polymers and organic materials. With the invention of more efficient mass analyzers, such as TOF, SIMS was used for more variety of applications, such as biomaterial research. Until recently, TOF-SIMS still suffered from the poor secondary ion production efficiency of the traditional gallium or indium ion beams, which limited the mass to low molecular weight species. Major breakthroughs have come more recently by the addition of gold cluster and bismuth ion guns as well as matrix enhanced SIMS (ME-SIMS). More detailed information on these can be found elsewhere[73]. These advances have expanded the application of TOF-SIMS to direct tissue profiling/imaging of biomolecules. Mouse brain images have been obtained showing images of ions from cholesterol, fatty acids, sulfatides, phosphatidylinositols, and triglycerides[74]. Muscle tissue from mice with muscular dystrophy has also been imaged[75]. At current, the TOF-

SIMS direct tissue technology is in the development phase, but holds potential to provide complimentary information to the current MALDI imaging technology.

In comparison to MALDI, the ionization process is not as soft and generates many fragment ions rather than intact molecules. The current secondary ion production is also lower than desired, there are not as many higher molecular weight molecules produced, which limits its mass range (up to 1000 Da). An advantage of this is the lack of need for matrices, whose ions tend to dominate the lower mass as well as create adducts in MALDI. This could allow TOF-SIMS to be complimentary to MALDI when it comes to the tissue analysis of lower molecular weight biomolecules such as cholesterol, phospholipids, fatty acids, etc. Another advantage is the potential nanometer lateral resolution for bioimaging in contrast to the micrometer lateral resolution in MALDI imaging. This technology, however, needs further development before it can successfully facilitate biomarker discovery for high-throughput clinical proteomics applications.

The strength of direct tissue analysis by mass spectrometry is the ability to obtain spatial information of biomolecules by directly analyzing tissue specimens. While this technology has been useful in obtaining biomarker information for cancer and other diseases, it has not been utilized for assessing molecular distributions in and around tumor margins. This would provide a further understanding of how the molecular characteristics around tumor margins differ from what can be observed histologically and how these characteristics change with higher stage tumors. As indicated, a preliminary study with one sample shed light on the potential of direct tissue analysis by MALDI-TOF MS to investigate these unknowns. A more in-depth investigation, using clear cell renal cell carcinoma, established the methodology to examine these unknowns and

illustrated the ability of direct tissue analysis by MALDI-TOF MS to facilitate such a study. These results will be described in Chapter IV.

Biocomputational Analyses

The large, complex datasets produced by clinical proteomic studies described above require robust bioinformatics tools to assess the patterns and decipher the proteins that are important in differentiating one disease state from another. For clinical proteomic biomarker discoveries, resulting MS data from both diseased and non-diseased samples are compared, using computer algorithms, to discern a pattern consisting of several discrete MS peaks (amplitudes increased or decreased) that are significantly different between the two groups of subjects. The ability to process and decipher this vast amount of proteomic data was initialized for genomic microarray data, which produces similarly complex sets of data. Many of these algorithms have since been adapted for proteomic dataset analyses. Numerous methods have been developed and are still being developed, but some scrutiny exists because each biostatistics method can produce different results. Thus, it is important to consider carefully the methods used, and perhaps apply more than one algorithm in order to be more confident of the final results. A more detailed discussion of data treatment and analysis will be found in Chapter III.

Clear Cell Renal Cell Carcinoma

In 2005, there were approximately 36,160 new cases and 12,660 deaths in the United States attributed to cancer in the kidney or renal pelvis[76]. Of these, approximately 75-80%[2, 77] are renal cell carcinoma (RCC), which includes a group of unique histological subtypes: clear cell (conventional) RCC, papillary (chromophil) RCC, chromophobe cell RCC, collecting duct RCC, medullary RCC, mucinous tubular and spindle cell RCC and other unclassified tumors[78]. The majority (80%) of RCC cases are classified as clear cell RCC (ccRCC)[77], called such because of the high content of lipids and cholesterol, creating a clear cytoplasm[78]. Clear cell RCC is also the most malignant of renal tumors, having the worst prognosis of all subtypes.

An Update on ccRCC

Genetics of ccRCC

Clear cell tumors occur in a hereditary form or a sporadic form. The inherited ccRCC is associated with the von Hippel-Lindau (VHL) syndrome[79]. Besides possible tumor development in the kidneys, individuals with VHL may also develop tumors in the cerebellum, spine, retina, inner ear, pancreas or adrenal glands[80], and are always those of the clear cell subtype[81], with affected individuals being at risk for up to 600 clear cell renal carcinomas per kidney[82]. Approximately 90% of sporadic ccRCC tumors involve the short arm of chromosome 3 (3p)[83-85].

Physical Attributes of ccRCC

In many cases ccRCC is clinically silent for years and discovered through symptoms from metastasis[86]; however, when symptoms are present, ccRCC may present itself with pain, as a palpable mass, or with hematuria. Grossly, the majority of ccRCC tumors are solitary and randomly distributed in the renal cortex. The mass is usually present as a rounded, lobulated, yellow mass that may protrude from the renal cortex. The size of the tumor can range from millimeters to weighing several kilograms. Although a yellowish color is typical, most tumors have a diversified coloration due to the presence of hemorrhage and necrosis. The majority of ccRCC masses may be well-confined from adjacent tissues or invade the adjacent renal tissue in an infiltrative manner. Most ccRCC tumors demonstrate a compact alveolar structure, containing numerous capillaries and thin-walled blood vessels in the surrounding stroma. This is a diagnostic feature commonly retained in the ccRCC metastases. Clear cell tumors occur with equal frequency in either kidney, with bilateral tumors existing in 0.5-3% of patients[2, 78, 86, 87].

Microscopically, ccRCC tumors have transparent cytoplasms, containing variable amounts of cholesterol, glycogen, neutral lipids, and phospholipids, with cells having well-defined cell membranes. Some cells will have a granular eosinophilic material around the nucleus and are characteristic of higher nuclear grade tumors. In clear cell tumors, the nuclei are round and oval and somewhat regular, but heterogeneity can exist within a single tumor, such as in ccRCC where sarcomatoid differentiation can occur, as discussed below[2, 78, 87].

Clinical Assessment of ccRCC

Preoperative assessment of RCC patients has three major aims: diagnosis of RCC, assessment of the tumor stage, and determination of details needed to plan the operation. Renal cell carcinoma is resistant to non-surgical methods, and needle biopsies of renal masses are prone to inaccuracy and so are not recommended for treatment decisions[88], thus, these tumors are primarily managed by surgery[89]. Once the diagnosis of RCC has been made, the staging is accomplished using radiologic assessment (MRI and multidetector computed tomography) of the local coverage, determination of whether or not there is venous or nodal involvement, and assessment of the lungs and liver for potential metastatic lesions[89].

The conventional and non-subjective diagnostic system used is the TNM staging system, which is the most important prognostic factor due to its positive correlation between the tumor at the time of diagnosis and its future prognosis. Although various schemes have been applied, the most widely used system is based on that proposed by the International Union Against Cancer (IUAC) and the American Joint Committee on Cancer (AJCC). The TNM (primary tumor, lymph node, metastases) system is based on anatomical examination and imaging, with the stages noted by pT, pN and pM. Many combinations of T, N, and M exist, but each combination can fit into one of four stage groups [Figure 5]. This system was established in 1978, but has undergone numerous modifications, with the current system being modified in 2002, and remaining under constant review[2].

Post surgical excision, the Fuhrman system[90] is conventionally used to classify tumors (fixed an paraffin-embedded) based on cellular morphology, including nuclear

size, contour, and the presence of nucleoli. Grade I tumors have small (~10 μm), uniform nuclei, with inconspicuous or no nucleoli visible[2, 78, 86]. Grade II tumors have larger nuclei (~15 μm) with variable sizes with some nucleoli visible at high magnification (400x). Grade III tumors present larger and more pleomorphic nuclei with prominent nucleoli visible at lower magnification (100x). Large, multilobulated nuclei are seen in Grade IV tumors, which sometimes demonstrate spindling and severe nuclear anaplasia, resembling a sarcoma, which is then referred to as the sarcomatoid variant of ccRCC. When tumor heterogeneity is present, Grade IV is always assigned[78, 86]. In contrast to the TNM staging, the histology grading system is much more subjective and tends to vary between pathologists, meaning that two pathologists may assign two different grades to the same tumor based on their individual assessments.

T (Primary tumor)

TX primary tumor cannot be assessed

T0 No evidence of primary tumor

T1a Tumor limited to kidney, ≤ 4 cm in greatest dimension

T1b Tumor limited to kidney, ≥ 4 cm but ≤ 7 cm in greatest dimension

T2 Tumor limited to kidney, ≥ 7 cm in greatest dimension

T3a Tumor invades adrenal gland or perinephric tissues but not beyond Gerota's fascia

T3b Tumor grossly extends into renal vein or vena cava below diaphragm

T3c Tumor grossly extends into vena cava above diaphragm

T4 Tumor invades beyond Gerota's fascia

N (Regional lymph nodes)

NX Regional lymph nodes cannot be assessed

N0 No regional lymph node metastasis

N1 Metastasis in a single regional lymph node

N2 Metastasis in more than one regional lymph node

M (Distant metastasis)

MX Distant metastasis cannot be assessed

M0 No distant metastasis

M1 Distant metastasis

I	T1	N0	M0
II	T2	N0	M0
III	T1	N1	M0
	T2	N1	M0
	T3	N0, N1	M0
	T4	N0, N1	M0
IV	Any T	N2	M0
	Any T	Any N	M1

Figure 5: 2002 TNM classification and stage groupings for clear cell renal cell carcinoma. (Adapted from Grignon et al[2])

The Need for ccRCC Tumor Markers

The determination of molecular markers that can be used to determine progression of a disease or its arrest, are essential to evaluate patients with RCC both to initiate new adjuvant treatments and to predict disease evolution. As mentioned, clinical staging is a valuable tool for ccRCC because of its high correlation with patient prognosis; however, no molecular markers are available to confirm the diagnosis or aid in further understanding of how this cancer spreads anatomically. Furthermore, new adjuvant therapies are greatly needed because ccRCC is virtually resistant to conventional chemotherapy or immunotherapy, thus known molecular markers that coincide with tumor stating could help determine which treatment would best suit individual needs based on the tumor characteristics, both anatomically and molecularly[91].

Over several decades, many possible prognostic factors have been studied but few of them have maintained an independent significance in terms of overall survival as assessed by multivariate analysis[92-94]. With this said, the clinical means used to assess these tumors remain to be stage, histological grade and performance status of the patient. There has yet to be identified any molecular or cytogenetic tumor marker that can help diagnose, manage or confirm ccRCC.

Gene Expression Profiling of RCC

Until recently, most molecular studies on RCC only classified renal cell tumors as one group from normal kidney but did not focus on genes specific to individual subtypes, which is crucial information for diagnosis, prognosis, and therapy of RCC, because each subtype has a different clinical outcome. Recently, Schuetz *et al.*[95] used oligonucleotide

microarrays to analyze 31 adult renal tumors consisting of ccRCC, papillary RCC, chromophobe RCC, oncocytoma, and angiomyolipoma tumors. Their results showed that ccRCC over-expressed 402 and under-expressed 220 genes. The over-expressed genes included mostly immune response genes, such as major histocompatibility complex II genes, and angiogenic factors. The clear cell subtype also expressed markers of proximal nephron epithelium, such as cubilin (*CBLN*) and megalin/low-density lipoprotein related protein 2 (*LRP2*), which were consistent with histogenetic models that relate ccRCC subtype to proximal nephron epithelium. Overexpression of angiogenic factors and receptors were also observed.

In a microarray analysis of 33 RCCs (clear cell and chromophobe) and 9 normal kidneys samples, Yao *et al*.[96] found 149 genes to be three fold higher in ccRCC than in chromophobe RCC and normal kidney. Out of those genes, adipose differentiation-related protein and nicotinamide *N*-methyltransferase were selected for further analysis and their overexpression was confirmed by real-time quantitative PCR and immunohistochemistry. The adipose differentiating-related protein is a lipid storage protein whose transcription is considered to be regulated by the von Hippel-Lindau/hypoxia-inducible factor pathway, suggesting the possibility that sustained upregulation of this gene following *VHL* inactivation is involved in the morphological appearance of ccRCC.

Gene expression profiling of ccRCC tumor tissue from 10 patients by Diegmann *et al.* identified *CD70* as a diagnostic biomarker for ccRCC, followed by confirmation using real-time RT-PCR and immunohistochemistry (IHC). Other genes found differentially expressed in their studies, between ccRCC tumor and normal kidney, that

were also identified by others include VWF, VEGF, VCAM1, PAI1, HEVIN, CCND1, CALB1, *CXCR4*, collagen types II and IV, integrin type alpha I, alpha 5 and beta 2, and FRA2, but half of these genes showed greater than twofold expression in only 50-70% of the patients studied.

In another study, Kosari and colleagues[77], were able to correlate survivin (BIRC5) with ccRCC aggressiveness by gene expression analyses. Survivin, a member of the inhibitor of apoptosis protein family, has been associated with aggressive behavior in tumors of the larynx[97], prostate[98], lung[99], and stomach[100]. Several genes showed increased expression in aggressive (high TNM stage) primary tumors (n = 9) as compared with nonaggressive (low TNM stage) tumors (n = 10) ccRCC. These included *IL8* (chemokine for inflammatory cells), *SAA2* (serum amyloid A2), and *CSK2* (CDC28 protein kinase regulatory subunit 2). Many genes were found to have decrease expression, including *NX17* (kidney specific membrane protein) and PLN (phospholambin). In this study, they used three different biostatistics algorithms and reported the best candidate biomarkers common to all three of the algorithms.

Proteome expression profiling

Proteins are encoded by genes, but their cellular production and activity is often controlled independently. Because proteins are the direct effectors and mediators that determine disease phenotypes, it is advantageous to take a proteomic approach to establish markers for diagnostics and prognostics. Numerous investigations have been undertaken to search for a protein marker for RCC, which would be useful, not only to use in conjunction with diagnosis and prognosis, but also as potential targets for therapy.

Most of these studies, however, are nonspecific, and investigate RCC as a group and not as individual subtypes. Another problem is the lack of sensitivity of some of these biomarkers to be used in the clinical setting such as for histopathological analysis by immunohistochemistry.

In a study, Sarto *et al.*[101] used 2D PAGE to study the normal and tumor kidney tissues from ten patients with RCC of different subtypes. For protein identification, the authors used one or more of the following techniques: gel comparison with other SWISS-2D PAGE reference maps, amino acid composition, *N*-terminal sequence analysis, and immunostaining. Their studies showed a loss of expression of four proteins occurring in RCC. Only two of them, ubiquinol cytochrome c reductase and mitochondrial NADH-ubiquinone oxidoreductase complex I, were successfully identified. Seliger and colleagues[102] used 2-DE followed by peptide mass fingerprinting (PMF) using MALDI-TOF MS of human cell lines from primary RCC lesions and normal renal epithelium to identify and validate the variable expression patterns of fatty acid binding proteins in RCC. While cell lines are valuable research systems, primary tissue samples are desired because they incorporate human heterogeneity into the study.

In a study with clear cell renal cell tumors, Hwa *et al.*[103] used 2D PAGE followed by MALDI-MS PMF analysis to obtain the proteome profiles of surgical specimens from ccRCC and surrounding normal kidney from a small set (n = 7) of patients. Two proteins, vimentin and alpha-1 antitrypsin precursor, were dominantly expressed in ccRCC. Proteins with repressed expression in ccRCC as compared to normal were aminoacylase-1, anoyl-CoA hydratase, aldehyde reductase, tropomyosin alpha-4 chain, agmatinase and ketohexokinase. The current number of reports applying whole proteome

analysis to clear cell and other renal tumors has been limited due to the lack of animal models for any subtype of RCC. Since primary tumors represent human and cellular environment heterogeneity, surgical tissue specimens are desired, yet it is difficult to obtain the quantity needed for statistical validation. Information on other RCC proteomic studies can be found elsewhere[2, 78, 87, 104, 105].

Despite the recent advances in RCC studies using integrated functional genomics and proteomics, the discovery of an early biomarker or set of biomarkers suited for use in the clinical setting for diagnosis or confirmation of the prognostic-like staging of any of the RCC subtypes has yet to transpire. Thus, more proteomic studies, with a larger set of samples, focusing on individual subtypes of RCC as well as addressing the molecular characteristics indicative of tumor stage are much needed[104]. This study aims at utilizing direct tissue analysis by MALDI-TOF MS to facilitate the discovery of biomarkers indicative of the clear cell subtype and its different tumor stages, which coincide with patient prognosis.

Problems with the Surgical Management of Tumors

One of the major issues in clinical oncology and pathology is ensuring complete tumor removal, which is essential for decreasing tumor recurrence rates[106]. Currently, histopathological assessment of hematoxylin and eosin stained sections of the resected tumor is the primary post-operative, and in some cases intraoperative, method for assessing surgical margin status. Depending on the type of tumor, the margins analyzed can be taken one to two centimeters away from the gross tumor margin. If a microscopic tumor is present at the surgical margins, the rate of local recurrence increases and the

survival rate decreases[107]. Incomplete removal significantly influences the likelihood of local recurrence in many cancers, such as lung[108], breast[109], soft tissue[110, 111], head and neck[107, 112], and brain[113]. Depending on the type of tissue, recurrence can occur a few months to a few years after removal of the primary tumor. Given that molecular alterations precede phenotypic changes, the environments outside of the histological tumor margin may not yet appear abnormal upon histological assessment, which renders conventional histological methods inadequate in being the only determinant in successful resection. With the lack of understanding on how seemingly normal cells become transformed into malignant cells outside the tumor margin and the inability of pathologists to visualize this transformation, molecular markers have the potential to change the way clinicians analyze margin status by providing an additional, and yet complimentary, way to identify aberrant tissue environments outside of the histological tumor margin[107].

Clinical Concern for RCC Margin Status

Due to the oncologic success of the radical nephrectomy procedure for renal cell carcinoma, it has become the standard treatment for RCC. However, for patients with RCC in a solitary functioning kidney, or those with RCC arising bilaterally (both kidneys), alternative approaches are needed. This was the initial reason for establishing the use of nephron sparing surgery (NSS) (also called partial nephrectomy), however, uses for this technique have broadened over time and now include avoiding complete removal of the kidney in patients without the aforementioned imperatives. These patients include individuals in whom the normal contralateral kidney could be under potential

future threat from hypertension, diabetes mellitus, or hereditary cancer like von Hippel-Lindau syndrome. Lastly, more kidney tumors are caught early on while they are small and radical nephrectomy represents over treatment. For these small tumors, NSS is also more appropriate because it can prevent potential problems such as hyperfiltration injury to the contralateral kidney[89].

Unlike radical nephrectomies, the oncologic effectiveness of NSS has been questioned, with the main concern being tumor involvement of the surgical margin. Surgeons have no true guidelines on how large of a surgical margin is needed to ensure complete removal of the tumor. While sparing as much of the kidney as possible, clinicians also need to ensure that the tumor is completely removed in order to eradicate the possibility of recurrence. Traditionally, it has been recommended that the tumor be resected with at least a 10 mm margin of normal parenchyma[114], however recent data provides evidence that a smaller margin might be adequate[115-117]. The optimal approach to ensure an uninvolved margin would be to perform an excision with a generous portion of normal tissue around the tumor, however, this would result in the loss of more healthy nephrons, which is not a realistic option for patients with solitary kidneys or those with renal insufficiency[118].

In-Situ Molecular Assessment of Tumor Margins by MALDI MS

Proteomic tools would be invaluable to discover how proteins change within tumor margins and the cellular characteristics in the environments that are just outside of the histologically-defined tumor boundary. Such studies would provide an estimate of how far outside the histologically-defined margin that cells are compromised. To

accomplish this task, proteomic methods must first be developed and applied to a system to determine the effectiveness of the methodology and how it can be applied to any system.

Summary and Research Objectives

Proteomics tools, such as 2D gel electrophoresis and LC-MS/MS strategies, have been successfully applied to facilitate the discovery of potential disease biomarkers; however, many of these tools are not very high-throughput and offer no information on the spatial distribution of biomolecules. The introduction of direct tissue analysis by MALDI-TOF MS for clinical proteomics has overcome high-throughput limitations while allowing for the collection of spatial information. Multiple sample analyses can be obtained in a day, which includes sample preparation and MS analysis. Direct tissue analysis by MALDI MS has proven useful for discovering protein patterns that distinguish between disease states, such as tumor and normal, as well as patterns that correlate with tumor stage and patient survival. The major strength of this technology, which is the ability to incorporate spatial information of biomolecules, has not been used to investigate one of the major problems facing clinicians and cancer patients, that is, how far outside of the histologically-defined tumor margin should surgeons resect to ensure complete tumor removal.

This research focuses on the application of profiling/imaging MALDI MS direct tissue analysis to the examination of tumor margins, using clear cell renal cell carcinoma, while also discovering molecular species that may distinguish between tumor stages and tumor grades. The goal of this study is to develop a methodology that can be applied to

any tumor margin system, with slight modifications. Clear cell renal cell carcinoma is not an attractive tumor for this study when considering its incidence and mortality as compared to other tumors; however, tumor margin status in ccRCC is now a concern due to the increasing demand for nephron sparing surgeries. Since the clinical management of this tumor is radical nephrectomy, ccRCC is an ideal system for this study due to the ease of acquiring the samples needed in contrast to other systems. The overall goal of this study was to determine the molecular characteristics of the cells outside of the tumor margin that appear histologically normal yet may actually be compromised. It is thought that there are molecular signatures and indicators of abnormal cellular development outside of the histological tumor margin, and that these molecular signatures can be correlated with tumor aggressiveness.

Using profiling/imaging MALDI MS, this research was conducted under the following objectives:

I. Determine if there are molecular indicators of aberrant cellular environments outside of the histological tumor border
II. Correlate molecular expression patterns with pathology-based tumor classification
III. Determine expression patterns of biomolecules that traverse the tumor margin into the adjacent normal tissue and correlate these with pathology-based tumor classification

Successful completion of these objectives will provide a methodology by which tumor margin characteristics can be investigated in other cancers. Identifying molecular patterns specific to ccRCC will enhance the ability to accurately diagnose patients by providing an additional molecular markers. Identifying characteristics of the cellular environment outside of the histologically-defined margin will not only provide new biological insights into the way margins are currently assessed, but provide a method by which the molecular events around the margin may be assessed in tumor types. In the long term, new molecular markers are needed to augment the current histopathological assessments of tumor margins to ensure complete tumor removal. These features will also allow for future, functional studies to better understand the mechanisms of local ccRCC tumor spread.

CHAPTER II

TISSUE PREPARATION AND MASS SPECTROMETRY ANALYSIS

Sample preparation procedures for direct tissue analysis by MALDI MS must be optimized for each tissue type to achieve sensitive and reproducible results required for protein profiling and subsequent statistical analyses. This chapter describes aspects of tissue sample preparation and MS analysis, including MALDI target plate selection, matrix solution deposition, and MS parameters to fine-tune for optimal signal. Lastly, this chapter includes a discussion on aspects of experimental design needed to minimize biases prior to statistical analysis.

Tissue Preparation

Tissue preparation procedures are important in experimental design. Proper handling of tissue will maintain the *in vivo* state of the sample, the three-dimensional shape of the sample and the spatial orientation of the biological compounds. Thorough forethought must be given to how the sample is collected, sectioned, and thaw-mounted onto a MALDI target plate, as well as what type of target plate is used and how the matrix is applied. Finally, the importance of consistency in preparation will be discussed.

Tissue Collection

Depending on the study, the tissue specimens collected for research come from animal models or humans and similar considerations apply to both, but the collection of

human specimens, which were used in this study, is more difficult to control. Many studies require interdepartmental collaborations, as well as, patient consent. The first priority of clinical staff is the well-being of the patient, but after tissues are taken for clinical diagnosis, an additional specimen is acquired for research purposes. When surgically excising and storing the specimen, it is important to maintain the three-dimensional shape of the tissue and minimize protein degradation. As the focus is on the well-being of the patient, the first priority of the operating room staff is to get the resected tissue to the pathologist for evaluation. Post-resection, the tissue will usually sit at room temperature for some time prior to piecing off the sample needed for pathological analysis. After a sample is set aside for pathology, it is kept on ice, and a piece of this sample is removed for the requesting researcher. The research specimen is then held on ice or in a refrigerator until it is snap-frozen in liquid nitrogen. Ideally the procurement process should take a maximum of 30 minutes, though few studies have examined how quickly the molecular content of surgical specimens degrades. Nonetheless, short procurement times will minimize protein degradation and maintain consistency in specimen collection that could otherwise be a source of variability in the final experimental results.

Sample freezing is accomplished by one of two methods. In the first method, the tissue is loosely wrapped in aluminum foil and slowly frozen by lowering the tissue into liquid nitrogen over a one-minute time period. The alternative approach is to create a floating aluminum foil boat for the sample to lie in, which will gently freeze it in approximately one minute. Rapidly immersing the tissue into the liquid nitrogen is not recommended because it often shocks the tissue, causing sample cracking and brittle

edges[119, 120]. Snap-freezing tissue specimens into liquid nitrogen helps preserve the shape and greatly slows biological degradation. Degradation is not a concern for DNA-based analysis, but is a great concern for RNA and protein analyses.

Once the sample is frozen, it can be stored at –80°C for years without significant degradation[119, 120]. The sample should be placed into a freezer-safe container large enough to contain the specimen without compromising its shape. Placing the sample into a conical tube is not recommended because it destroys the tissue shape. If the sample is previously wrapped in aluminum foil, the wrapped specimen may be placed into a large container for labeling. Alternatively, the sample may be removed from the aluminum foil boat and placed into the container.

The sample collection process described above is ideal, but not always followed. Often, the collection is out of the researcher's hands and left solely to the tissue repository or collection facility. It was discovered in this study that many institutions offering to collect specimens for research opportunities are not properly designed to achieve procurement time under one to two hours, while others routinely procure samples within 30 minutes. This inconsistency indicates a need for these facilities to have uniform standard operating procedures. Consistency in collection procedures will ultimately minimize variability for RNA and proteomic studies. It is ultimately the responsibility of the repositories to create a general guideline for tissue collection, freezing, and storage prior to handing the sample over to the researcher[121]. Specimens used in this study were frozen at -80°C within 30 minutes of surgical resection.

Tissue Sectioning

For direct tissue analysis by MALDI MS, frozen tissues are sectioned using a cryostat for subsequent mounting onto MALDI target plates. Cryostats allow for accurate sectioning of samples at a given thickness while maintaining the spatial orientation of the tissue. Samples are attached to a movable holding stage and sliced with a stationary stainless steel microtome blade. The sample stage and the cryostat chamber temperature are generally held between –15ºC and –25ºC, depending on the tissue type. Fatty tissues require lower temperatures, in order to maintain spatial integrity. The temperature settings used for ccRCC samples in this study were –16ºC for the sample holder and –18ºC for the chamber temperature.

Sectioning may be performed in two ways: manual rotation of the sample holder wheel, or by automated, motor driven movement. Manual rotation of the wheel may introduce variability in the section due to variable force and speed used to turn the wheel. Motor-driven sectioning, however, introduces a consistent turning of the wheel. This allows for more reproducible sectioning, which eliminates additional variability in the overall experimental design. Motor-driven sectioning was performed in this study except in cases where tissue specimens and/or the cryostat anti-roll bar were difficult to work with.

Traditional microtome sectioning in pathology labs for histological staining uses an embedding medium, such as paraffin, which maintains the sample at room temperature and is ideal for preserving the tissue morphology for histological evaluation. However, the use of such polymers is not compatible with MALDI MS, as they cause ion suppression, thus, fresh frozen samples are required for MALDI MS analysis. Optimal

cutting temperature polymer (OCT) is the medium used for cryo-sectioning frozen samples for histological or other types of evaluation. OCT is also a polymer and, as mentioned, will suppress ion formation in the MALDI source[122]. It is therefore recommended that the sample be carefully mounted onto the sample holder so that the OCT touches only the bottom portion of the sample, avoiding any contact with the surface being sectioned for MALDI analysis [refer to Figure 12-1a].

The section thickness depends on the sample and the methods used. Thicker sections result in less tearing, making them easier to manipulate onto the MALDI plate. If the section is too thick, it will take longer to dry, promoting potential protein migration, and may not be as electrically conductive once in the mass spectrometer. Thinner sections, in contrast, are more difficult to maneuver and may tear easily; therefore, for MALDI applications, tissues are optimally cut between 10 and 30 μm. With the average mammalian cell diameter between 10 and 20 μm, the majority of cells in the section are sliced open, exposing the intracellular components[122]. A thickness of 12 μm was used for this study.

It has been suggested that the blade used in the sectioning process may smear proteins across the tissue surface. Although sectioning of samples has been performed routinely in immunohistochemistry and autoradiography studies with little to no indication of protein delocalization, the sectioning process was validated for this study. To ensure that the blade did not cause protein delocalization for this study, one section of a ccRCC sample that contained tumor and adjacent normal was sectioned and thaw-mounted onto a gold-coated MALDI plate. The tissue block was then rotated 90º or 180º and another section was obtained and mounted onto the same MALDI plate. An array of

matrix spots (method described below) was deposited on both tissues and analyzed by MALDI-TOF MS. Ion-density maps were created from the spectral information obtained at each matrix coordinate to demonstrate the localization of proteins. This experiment was repeated in triplicate on three different days, and a representative result is shown in Figure 6. Resulting images show no evidence of protein delocalization due to the cryostat blade.

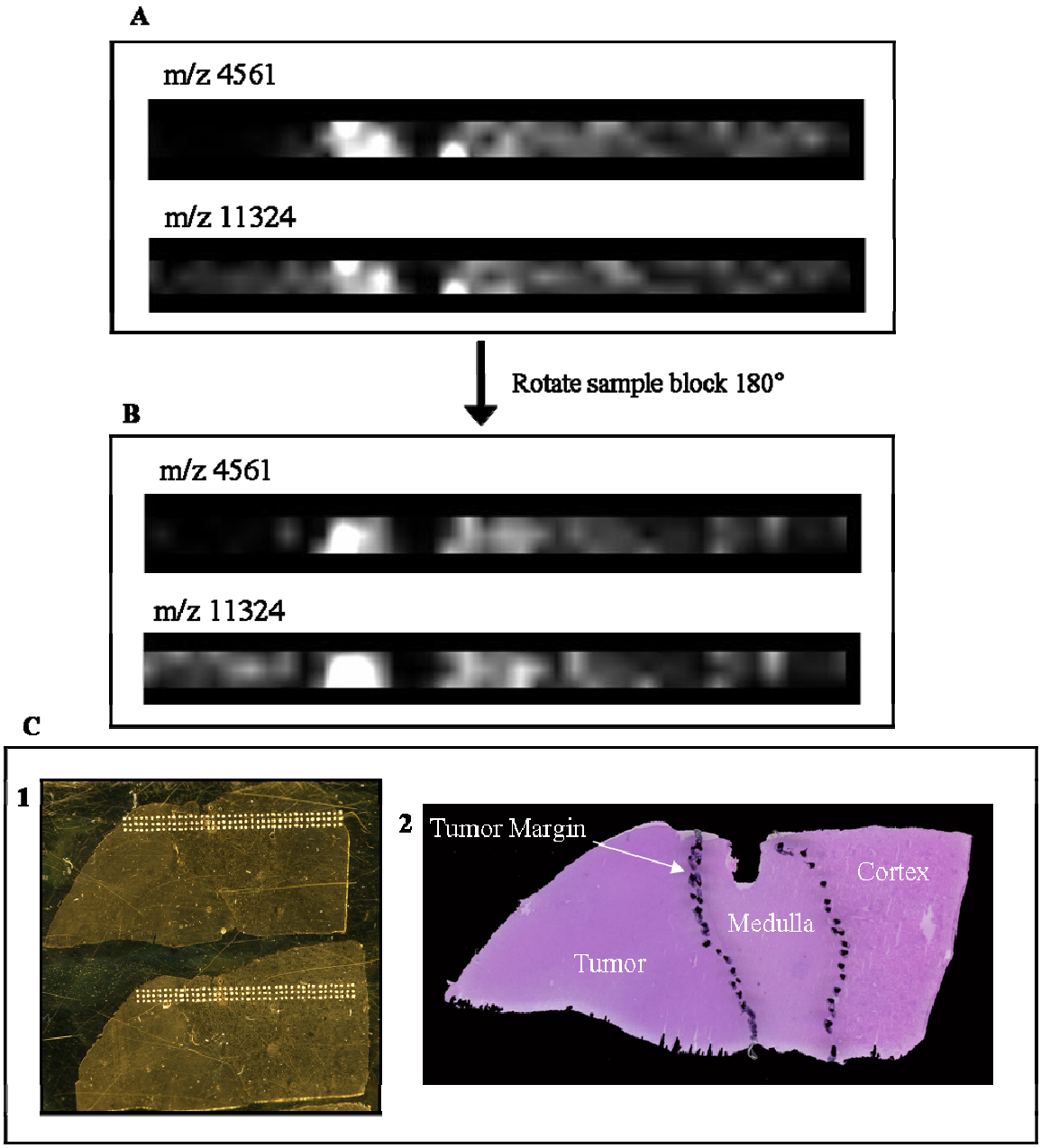


Figure 6: Cutting Validation. A.) Representative ion density maps from reference section. B.) Representative ion density maps from the section that was cut 180° to the section imaged in A. C.) C1 is an optical image of the sections on the target plate prior to imaging; the second section was rotated 180° with reference to the top section. There is approximately a 144 µm distance between the two sections. C2 is a representative, histologically-marked H&E section of the specimen analyzed. This figure demonstrates that there was no protein delocalization observed due to the cryo-sectioning process.

Tissue Transfer and Mounting onto a MALDI Plate

There are two methods for the transfer and thaw mounting of tissues onto MALDI plates. In the first method, the room temperature MALDI plate is placed over the frozen section, which adheres to the plate as it thaws. In the second method, the frozen section is cautiously picked up with small paint brushes and placed on top of the MALDI plate, which has been held at the chamber temperature prior to thaw-mounting (see below). The former method is not ideal. First, it is difficult to control the placement of tissue onto the plate because most MALDI plates are opaque (gold-coated or stainless steel), although recent plates include ITO-coated glass. Secondly, this method leaves behind ice crystals on the cryostat stage that may result in the loss of water-soluble proteins[123]. This method is also difficult to control, rendering it less reproducible than the second option.

The second method, in which the frozen section is transferred to the plate with small paint brushes before thaw-mounting, does not result in protein loss; however, it requires care and skill to maneuver the sections without tearing them. Prior to thaw-mounting, tissue must be transferred in a way that it retains its shape, without tears and rolled-up edges. When these deformities occur, they can potentially mask important tissue-specific mass spectra, possibly affecting the final analysis. Distortions can also make it difficult to compare to subsequent sections placed on glass microscope slides for histology. After mounting, the plate and tissue are quickly warmed together with the finger or palm of one's hand on the back side of the plate to thaw mount the tissue onto the plate. This tissue-transfer method was applied to the ccRCC biopsies in this study.

Choosing a MALDI Target Plate

There are two types of MALDI plates: metal (usually stainless steal or gold-coated) and glass (e.g. ITO-coated)[123]. Recent studies have illustrated the affects of sample thickness and preparation methods on the ionization efficiency of metal substrates[124]; however, stainless steel plates (matte and polished) are used less frequently in direct tissue analysis by MALDI because the poor contrast between the plate and the tissue makes it difficult to see with the MALDI instrument cameras and in digital snapshots[123]. The polished gold-coated plates provide ample contrast for the desired digital snapshots and visualization, as well as ability to visualize differences in tissue morphology like tumor and non-tumor regions.

A disadvantage to working with the metal plates is the need for subsequent sections to be collected on a microscope slide for histological staining in order to correlate cellular morphology with resulting MALDI spectra. Adding a conductive surface, such as ITO, to a glass slide makes it compatible with the MALDI process. With these plates, there is an option of using a MALDI-compatible nuclear stain, which allows for histology and MALDI analysis on the same section[123]. Despite the reported compatibility, it cannot be ignored that these stains induce noticeable signal differences compared to fresh frozen tissues, so the most optimal analysis is the native section, without addition of histological stains[125].

An additional concern when choosing a plate is its matrix spotting performance. With different coating properties of the ITO-coated and gold-coated MALDI plates, the matrix spotting performance is likely to be different with each surface and different tissue types. Each plate type was examined for renal tumor sections containing adjacent normal

renal tissue. The two types of MALDI plates containing the renal tissue of interest were robotically spotted[126] (discussed below) at a 400 μm spot-to-spot distance. The spots on the ITO-coated plate were not homogeneous and were placed in a disordered array with spots running into each other. This observation could be due to an extended drying time of spots on glass as compared to metal plates[126]. The seeding technique[126] applied prior to spotting on the ITO-coated slides slightly improved spotting; however, the gold-coated plates provided more homogeneous crystals in an ordered array without the need for seeding [Figure 7], which can introduce variability in the sample preparation process. The gold-coated MALDI plates were chosen for this study; therefore subsequent sections were obtained in the cryostat for conventional histological staining (hematoxylin and eosin) on microscope slides.

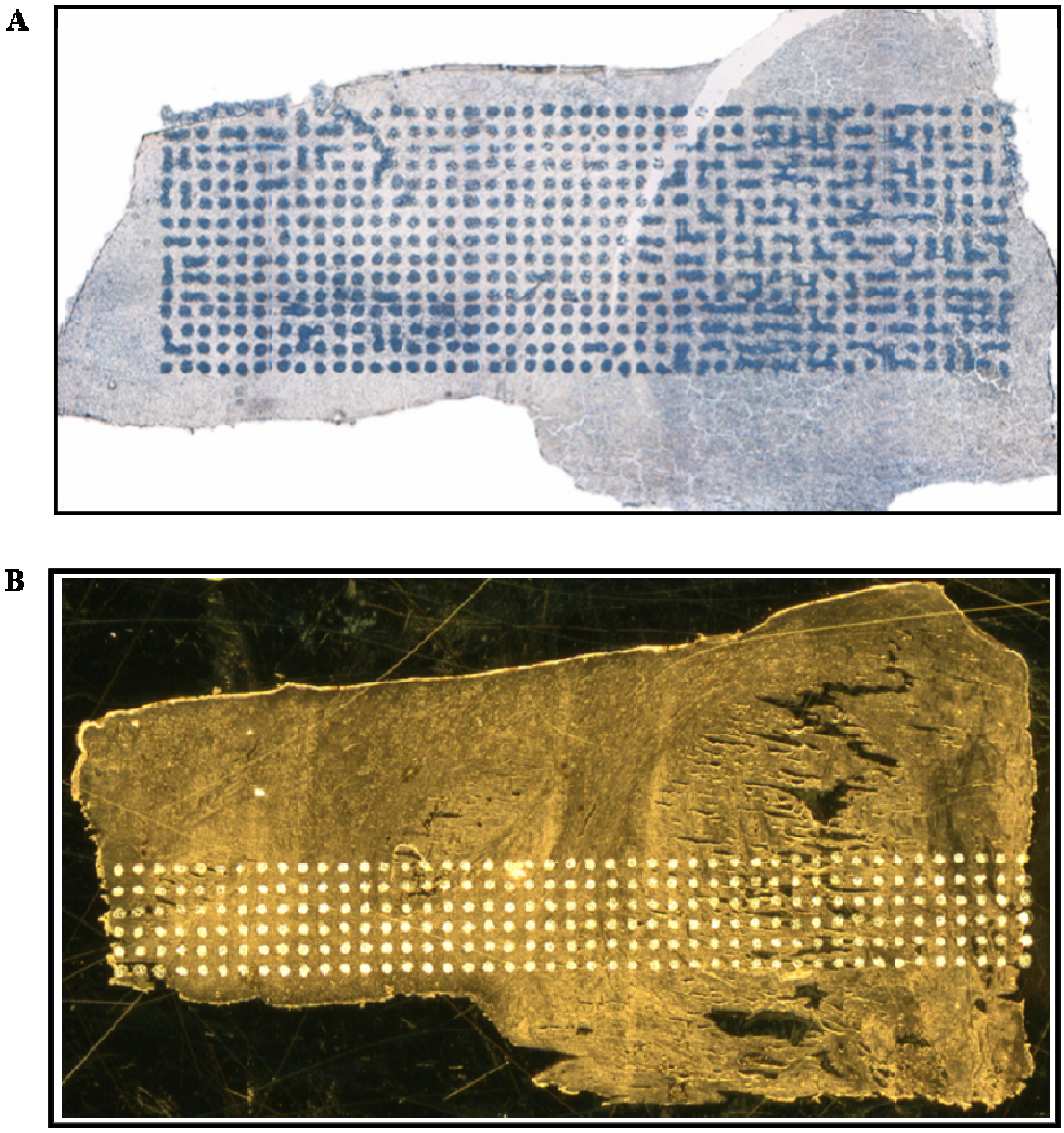


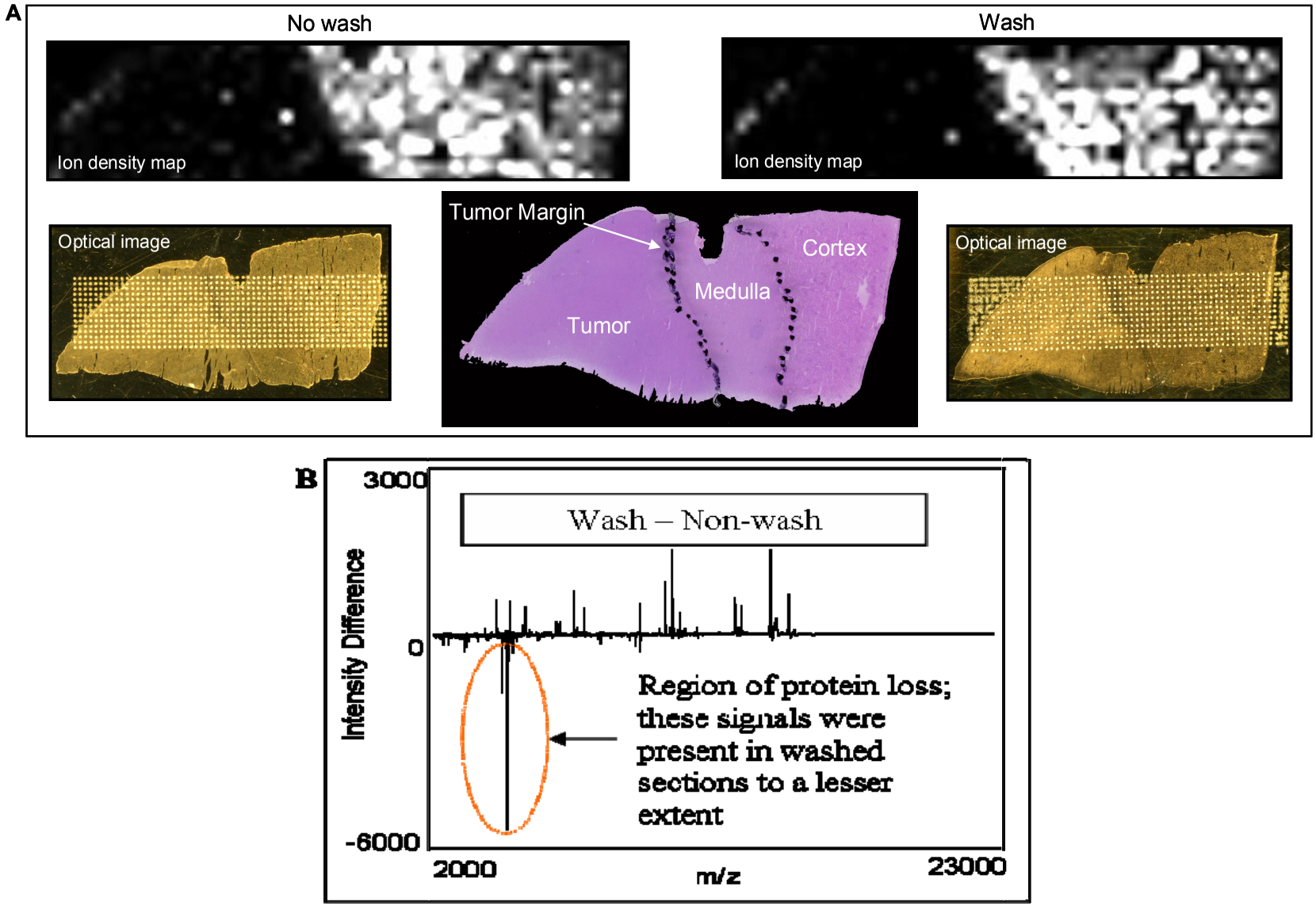
Figure 7: ITO-coated MALDI plate versus gold-coated MALDI plate. A.) Optical image of a spotted section on an ITO-coated glass plate. B.) Optical image of a spotted section on a gold-coated stainless steel-plate. This figure illustrates the superiority of the gold-coated plates over the ITO-coated glass plates for optimal matrix spotting onto ccRCC samples, even when the seeding technique is applied for matrix deposition as in A.

Tissue Washing

Depending on the tissue type and biomolecules of interest, it is often advantageous to wash the tissue sections prior to matrix deposition to remove salts and other contaminants, as well as, excess hemoglobin from the tissue surface that can cause a high spectral baseline and signal suppression. Although sinapinic acid (SA), the matrix of choice for protein analysis, is tolerant of salts and other contaminants[127, 128], the removal of these via washing does enhance signal quality[122]. The conventional tissue washing method involves two separate 70% ethanol (200-proof or reagent grade) emersions for 20-30s followed by 10-15s in 95% 200-proof ethanol (or 100% if using reagent grade ethanol). Using 100% 200-proof ethanol may over-dehydrate the tissue. Since ethanol is a known tissue fixative often used in histology[129], delocalization of proteins is not likely to be significant; however, proteins soluble in these aqueous solutions may be removed during the washing step. Previous studies have shown that there is minimal protein loss during this process, but different tissues may behave differently, so the washing procedure should be confirmed for each tissue type[122, 123, 125].

To validate the tissue washing step, two sections of a ccRCC sample that contained tumor and adjacent normal were sectioned and thaw-mounted onto a gold-coated MALDI plate. One section was washed as described above (2 x 70% ethanol for 20s, 1 x 95% ethanol for 10s) and the other was not washed. To determine the extent of protein loss in the washing step, the ethanol solutions were combined after washing and speed vacuumed to dryness. The solution was reconstituted in acetonitrile and analyzed by MALDI-TOF MS. An array of matrix spots was deposited on both the washed and non-washed tissues and analyzed by MALDI-TOF MS. Images were assembled and ion-

density maps were created to demonstrate the localization of remaining proteins. The average spectra from the washed and non-washed tissue were compared for overall signal enhancement or loss. The washed tissue showed more overall signal enhancement and better spectral quality than the unwashed tissue. The spectra acquired from the ethanol wash solutions were compared to the averaged spectra from both washed and non-washed tissues. This experiment was repeated in triplicate, on three separate days, and results from a representative experiment are shown in Figure 8. Results illustrate the following: 1.) the proteins remaining after washing show no evidence of protein delocalization due to the washing procedure; 2.) no significant protein loss occurred, other than a few low m/z signals; 3.) washing did enhance MS signal. Overall, there was some loss of protein signal in the low m/z region due to washing, but the overall enhancement of signal due to washing outweighed the few protein losses.


A
No wash
Wash
Ion density map
Ion density map
Optical image
Tumor Margin
Cortex
Medulla
Tumor
Optical image
B
3000
Wash – Non-wash
Intensity Difference
0
-6000
Region of protein loss; these signals were present in washed sections to a lesser extent
2000
m/z
23000

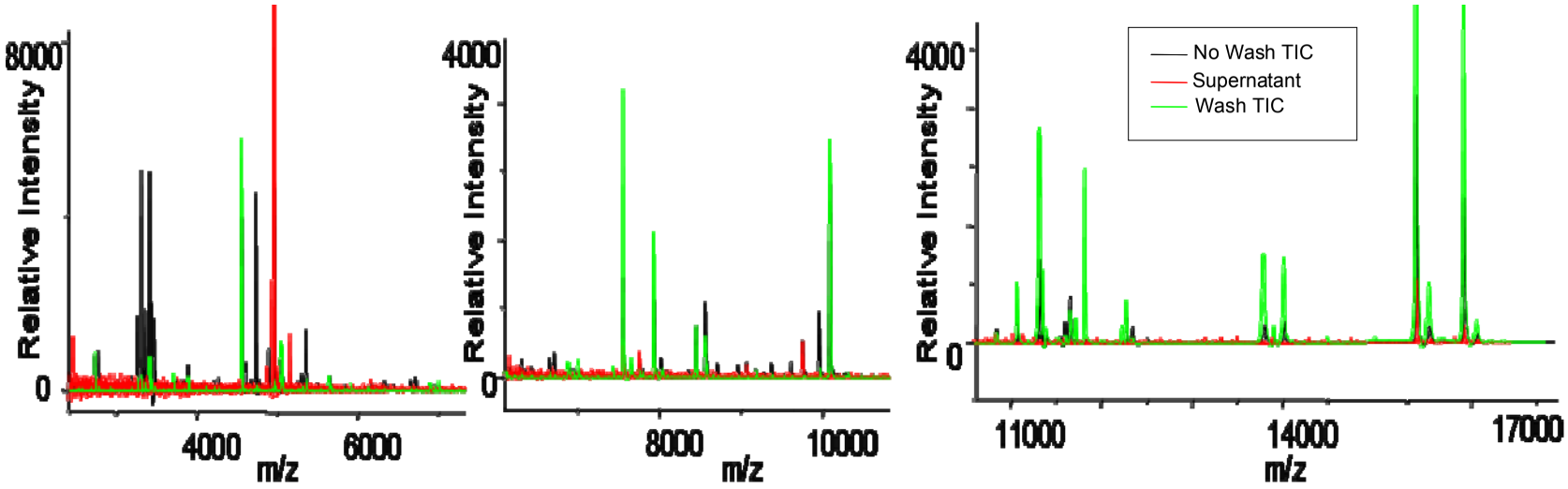


Figure 8: Validation of Tissue Washing. (A) A representative ion density map of protein localization in the washed and non-washed tissues. The optical image, as shown at the bottom of A, describes what is seen in the ion image. Comparison of protein localization in the two images shows no sign of delocalization at this image resolution (400 μm). (B) A direct comparison of the TIC from non-washed tissue to washed tissue is seen in the spectral subtraction of non-washed from washed tissue, which shows that the only significant loss was in the low mass range and the signal from the washed sections were significantly enhanced. (C) An overlay of the total ion current (TIC) from the non-washed tissue and washed tissue as well as the averaged spectrum from the ethanol wash supernatant collected from the washed tissue section. Each of the three spectra was baseline subtracted and normalized prior to the overlay; each spectrum window shown is a zoomed-in region of the total mass range collected (m/z 2000-25000). Results show that the ethanol washing steps result in some signal loss, but a significant overall enhancement of the MS signal.

Matrix Solution

Another important aspect of sample preparation is the matrix concentration and solvent composition. Crystal coverage and crystal homogeneity of individual spots are influenced by these factors. As previously mentioned, SA is the optimal matrix for protein/tissue applications. Previous studies have also shown that higher matrix concentrations result in higher quality protein spectra with a saturated solution (25-30 mg/mL) yielding the highest number of ion signals with best resolution compared to lower SA concentrations[122].

While there are many options for the matrix/solvent combination, the goal of the matrix solution is to solubilize the analyte of interest, or in this case, to extract as many proteins as possible from their native microenvironment within the tissue. Solvent conditions will then vary depending on the analyte of interest or the tissue type being analyzed. Various solvent/water compositions have been examined, which showed that 50% acetonitrile/water or 50% ethanol/water compositions were optimal. Both of these options consistently show no overall improvement over the other, so acetonitrile is the conventional organic solvent used in direct tissue analysis by MALDI MS[122]. Signal quality from ccRCC tissues was examined for varying acetonitrile concentrations (50-70%), but none showed significant improvement over the 50% composition (data not shown).

As the addition of trifluoroacetic acid (TFA) to the matrix solution aids in the ionization process, its concentration may affect the protein profile. A TFA concentration of 0.1-1.0% is optimal for MS analysis of tissue, with the 50% ACN/0.1% TFA composition most widely used[130]. For the ccRCC samples, various TFA concentrations

were examined in a range of ACN concentrations, and none showed overall improvement over the 50 % ACN/ 0.1% TFA solution (data not shown).

Matrix Application on Tissue

Manual Deposition Techniques

Prior to analysis by MS, matrix must be applied onto the tissue surface in a reproducible manner in order to achieve homogeneous crystallization. For direct tissue MS applications, there are two conventional strategies used to apply matrix to the tissue surface: depositing small droplets of matrix on specific regions of the tissue, and coating the entire tissue surface with matrix solution [see Figure 3]. The latter mentioned method is best applied when high-resolution images are desired. Descriptions of this method can be found elsewhere[122]. In the former method, small droplets of matrix (100-250 nL) are manually deposited, with a low-volume pipette, on discrete morphological regions of the tissue, from which protein and/or peptide signals in that defined region are to be obtained. Placement of matrix spots can be obtained via pathology-guided assistance from an adjacent, histology stained tissue, or from the tissue itself employing the MALDI-friendly dyes on ITO-coated plates. This method allows for a quick assessment of the protein content within a tissue section for further facilitating a comparison between different tissue states, such as disease vs. normal or treated vs. untreated.

Aside from being low-throughput and having poor reproducibility, another important drawback with manual matrix deposition is the matrix spot size, generally 1 mm in diameter. Given that the average cell diameter is about 10 μm, depending on the

cell cycle, a 1 mm diameter matrix spot will limit the specificity obtained in profiling experiments. With smaller diameter spots, it is possible to obtain more cellular specificity within the micro-environment of a morphological region. Attempting to address this need is the utilization of syringe pumps, to which a small capillary is connected for the placement of smaller, more reproducibly placed matrix droplets. This approach provides more specificity and reproducibility than manual deposition, but more optimal approaches, in terms of matrix spot size and spot-placement reproducibility, have been recently attained using robotic devices[126, 131].

Robotic Deposition Technologies

Although many automated devices are capable of generating and accurately depositing droplets onto a surface with precision, few are applicable for direct tissue analysis. Contact deposition methods that require a pin or capillary to facilitate reagent transfer through physical contact with the sample are often not reproducible and the physical contact with the tissue surface may cause cross-contamination and destruction of tissue integrity. Non-contact printing devices, however, hold promise for direct tissue applications. These devices include piezoelectric, thermal ink jet, solenoid valve, photolithography and pulsed field ejectors. With these, cross-contamination and tissue destruction are not a concern because there is no contact with the tissue.

Two non-contact robotic multi-spotters, designed for high-throughput and reproducible sample preparations for microarray and MS applications, were available for this study that have been applied to direct tissue analysis: an acoustic reagent multi-spotter (ARM)[126], custom-made at Labcyte Inc. (Sunnyvale, CA) and a chemical inkjet

printer (CHIP)[131, 132] by Shimadzu Biotech (Kyoto, Japan) [Figure 9]. In brief, the ARM technology, which was used for this study, deposits reagents without the use of a capillary or nozzle. Instead, it utilizes short pulses of focused acoustic energy directed onto a liquid surface from under an open reservoir containing the matrix solution.

To operate the instrument, the MALDI plate, facing downward, is first positioned in the inspection position on an X-Y translational stage, where a color CCD telescope captures digital images of the tissue sections and then displays them in the control software on the computer monitor. Matrix deposition locations are chosen by clicking the desired pixel in the captured image or by clicking and dragging a large rectangular array of spots at the desired spot-to-spot distance. Matrix spot coordinates and parameters are saved into a file and the MALDI plate is then transferred to a second holding position on the X-Y translational stage, facing downward, into the printing position located above the reservoir, which is above the ejector [Figure 9]. Prior to deposition, the ejector must be tuned for the minimum rf energy needed to produce a stable droplet. With the matrix solution in the reservoir, the rf energy is manually adjusted until the threshold energy is reached. This occurs when the rf amplitude is sufficient to cause the point of the liquid cone to break away as a spherical droplet. A transparent lid, containing a 400 μm diameter hole in the center for droplets to pass through, is then placed on top of the reservoir to minimize solvent evaporation, which will adversely affect the tuning. It was observed in both the previous study[126] and in this study that the tuning will last for about 120,000-150,000 shots, or approximately 5-6 hours of continuous spotting. More detailed information on the ARM can be found elsewhere[126].

The CHIP multi-spotter is a microjet device [Figure 9], which utilizes piezoelectric drop-on-demand-type ink-jet technology for rapid liquid microdispensing[131, 132]. Most recently, this technology has been used to deposit matrix arrays on tissue surfaces for direct tissue analysis by MALDI-TOF MS[133]. The CHIP instrument is designed with piezoelectric microjets, four total, connected to individual reagent reservoirs and an air pressure/vacuum control unit for fluid stabilization within each of the microjet devices. Fluid dispensing and motion of the MALDI target plate (positioned face-up below the microjets) in the X-Y translational directions are controlled by integrated software and controller boxes. Two cameras allow for visualization of droplet formation and a top view of the MALDI target plate, respectively. Image-capture software is used to define X and Y coordinates for automated spot delivery. Tuning is performed with the software to produce the desired droplet as it is microdispensed from one of the glass capillary piezoelectric devices. More information on this technology can be found elsewhere[131].

There are advantages to using the ARM or the CHIP. Both are non-contact spotting devices, which eliminate cross-contamination within an experiment or destruction of the tissue itself. In terms of spot-to-spot resolution, the lower limit of both instruments is around 200-230 μm. Both instruments can operate over several hours without operator interference. The CHIP spotter was not tested for this analysis, due to availability. The ARM spotter was originally used for this study and was used for all subsequent experiments to maintain consistency, although either spotter would have been appropriate.

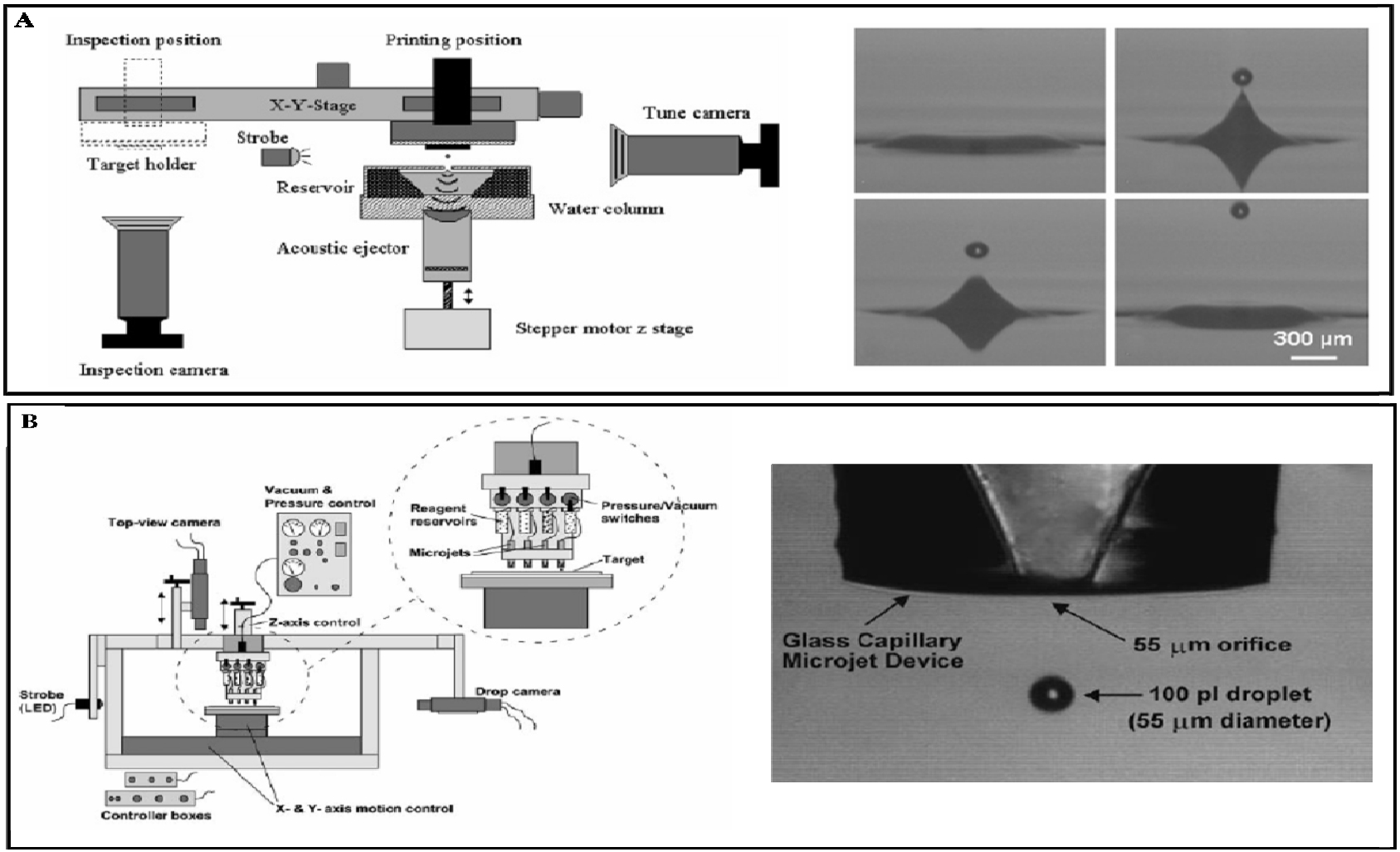


Figure 9: Schematic representation of robotic spotting devices. (A) ARM and acoustic-generated droplet from the ARM (Reprinted, with permission from reference [126]. Copyright 2006 American Chemical Society.) (B) CHIP and the piezoelectric-generated droplet. (Adapted from reference [131].)

Optimizing Parameters for ARM Matrix Deposition

Optimization of sample preparation procedures prior to and during the matrix deposition using the ARM has been examined, but must be evaluated for each matrix solution composition. It has been determined that as the percent organic (ACN) in the matrix solution decreases, the droplet volume increases, with 50% organic giving droplet volumes of 122 +/- 16 pL and a diameter of 61.6 μm. The ejection frequency and the number of successive droplets, however, also affect this diameter. Matrix spots increase in diameter with increasing number of droplets. Lower ejection frequencies appear to have less impact on spot diameters than higher frequencies. It was also determined that the density of matrix crystals and the number of MALDI spectra that can be obtained before analyte signal is depleted also correlates with the amount of matrix material deposited. With increasing droplets per spot, the average diameter increases from 213 μm for 9 drops to 291 μm for 60 drops (drops per spot per pass). In general, the optimum parameters for profiling experiments that were applied to this study are as follows: 10 Hz spotting frequency, 13 droplets/spot with 6 passes (totaling 78 drops/spot)[126] [Figure 10].

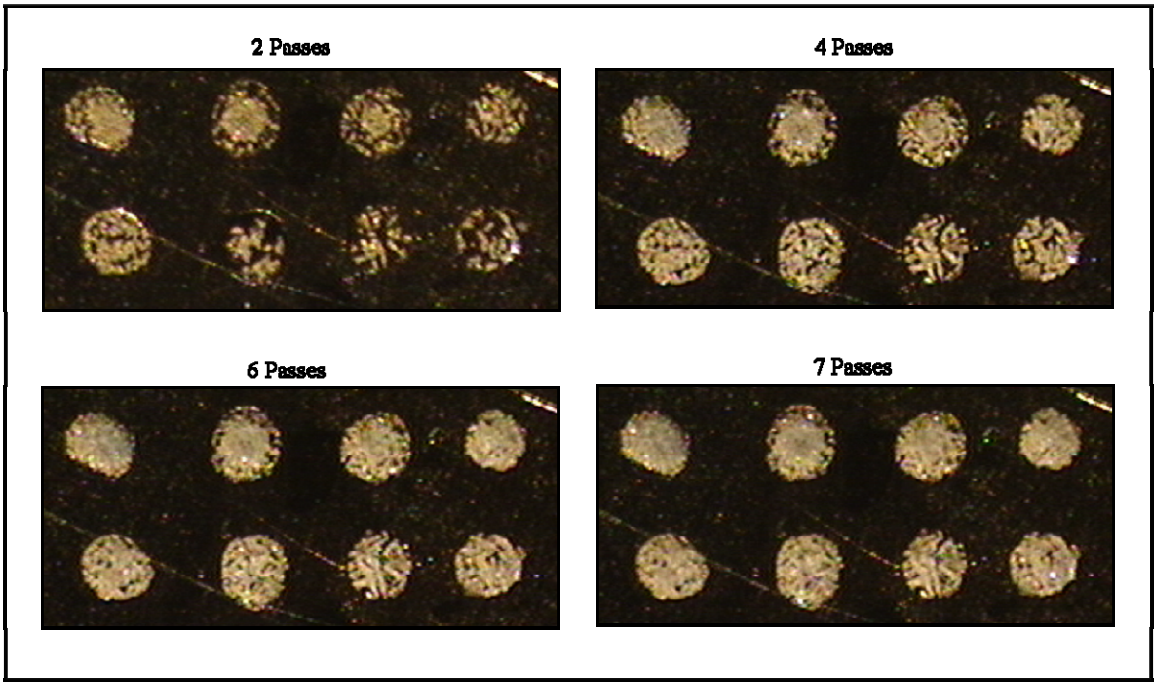


Figure 10: ARM matrix deposition. Each image represents matrix spots obtained at spotting frequency of 10 Hz at 13 droplets/spot. This figure illustrates how each pass of matrix deposition affects the uniformity and homogeneity of crystal coverage between spots and within a spot. Six passes result in the optimal spotting performance; as illustrated, few improvements are observed beyond 6 passes.

Automated MS Profiling from ARM Spotted Arrays

Automatic matrix printing followed by automated MS analysis offers high-throughput capabilities for rapid matrix deposition and data acquisition. Prior to MS analysis, a registration procedure is required that converts the spot coordinates from the tissue to the mass spectrometer. After spotting, the MALDI target plate containing the spotted tissue is scanned to obtain a digital image. The image is opened in Image-Pro (Media Cybernetics, Inc., Silver Spring, MD) or Adobe Photoshop (Adobe Systems Inc., San Jose, CA), where a colored mark is manually placed in the center of each matrix spot. The relative center coordinates of each spot are obtained using Image-Pro software. This information is used to create a custom geometry plate file required by the instrument for automated acquisition. The geometry file can be created manually or with a custom program. After loading the geometry file into the instrument control software and the plate into the instrument, a registration procedure teaches the instrument which matrix spot on the tissue corresponds to a specific spot in the geometry file.

Prior to spectral analysis, several automated mass spectrometer acquisition parameters must be defined. These include the number of laser shots to average per spot and any acceptance criteria, based on signal intensity that can be used to avoid collecting spectra of low quality. Matrix spots should be sampled with minimal cross-sampling within a spot and no cross-sampling of neighboring spots. With a laser diameter of 70-100 μm, the matrix spots are generally large enough to allow the collection of 200 to 600 mass spectra from several discrete locations within each spot without overlap. For this experiment, 400 shots per spot were averaged, and acceptance criteria for each spectrum were set, including S/N (at a ratio of 10) and resolution (no minimum was set). These

criteria should be set at a minimum to avoid introducing operator bias that may mask differences due to true biological changes.

MALDI Mass Spectrometric Analysis

Considerations for Optimizing Instrument Parameters

Prior to collecting spectra, it is essential to consider and eliminate many of the challenges brought about by the desorption of proteins from a complex mixture like tissue. Although many of the compounds such as salts (which provide Na^+ and K^+) and lipids that adversely affect the MS analysis should be washed off in the ethanol-washing step, there remain other proteins and biological compounds within the sample that will adversely affect signal. Although ion suppression and adduct formation will always be present, the goal of the MS set-up is to optimize parameters in a way to achieve minimum interference and achieve reproducible results. Instrument parameters are difficult to optimize for all masses across a large mass range, such as mass-to-charge (m/z) 2,000 to 70,000, and can compromise sensitivity and mass accuracy in some regions within this range. However, it is possible to obtain high-quality mass spectra by effectively optimizing the instrument parameters.

Optimizing MS instrument parameters

- Mass range

Beyond the sample preparation steps, further signal optimization must be done with the instrument parameters. As previously mentioned, it is difficult to optimize

parameters for a wide mass range; however, depending on the tissue type, acquiring such a wide mass range may not be necessary. In renal samples, few signals were observed above m/z 25,000, thus the mass range sampled was m/z 2,000-25,000.

- Laser, acceleration, and delayed extraction parameters

The overall goal to achieving quality spectra is to enhance sensitivity and signal-to-noise without a major sacrifice in resolution, while maintaining a relatively flat baseline. For any given mass range, on and off tissue, there is a general set of optimal values for each parameter to obtain these qualities. In most cases, these values will need to be fine-tuned, depending on the tissue type. After choosing the mass range, the instrument parameters to optimize for spectral quality on the instrument used in this study are as follows: laser attenuator (offset and range), laser focus (range and value), laser frequency, laser power, delayed extraction time (ns), ion source voltages 1 (acceleration voltage) and 2 (time lag focusing). Important aspects and functions of the more crucial parameters are described below.

When optimizing laser settings, the best spectral quality is often achieved when the laser power is set just above the ionization threshold level of the matrix. Excessive power will result in saturated peaks with poor resolution and high sample consumption. The amount of laser energy needed for the generation of MALDI ions depends on: the matrix, which has a well-defined energy threshold; the susceptibility of the analyte to the MALDI process; the matrix-to-analyte ratio; and the preparation technique used. Additionally, the number of laser shots averaged per spectrum for direct tissue profiling ranges from 250-1000, depending on the spectral quality, diameter of the matrix spot, and possible time constraints.

After ions are formed in the desorption process, they are accelerated into the mass analyzer by an acceleration voltage. The acceleration voltage (IS1) is an important parameter for resolution. The absolute difference in energy, and consequently in velocity and flight time, observed for ions of differing masses increases with rising voltage applied for initial acceleration of the ions formed during the desorption/ionization process. While these values are preset (20-25 kV), they can be slightly adjusted, while detector settings are optimized for specific acceleration voltages.

Ions of the same mass formed in the MALDI process have a range of translational kinetic energies, which contributes to peak broadening and loss of resolution[47, 134]. In a constant potential field, when molecular ions are formed, they are affected by the electrostatic field of the ion source and progressively accelerated. In doing so, they undergo multiple collisions with residual neutral matrix components of the desorption plume, which further widens their overall energy distribution, resulting in peak broadening. The initial velocity spread can be compensated for with delayed extraction (time-lag focusing). This requires that the source be comprised of two electrostatically independent stages. In principle, the ion plume formed during the MALDI process is allowed to remain in a field free environment, with equal potentials on the sample plate and first electrode, for a period of time (nanoseconds) before being pulsed out. At the end of a given delay time specified, the ions are pulsed out of the first stage of the source by either dropping the potential of the first electrode, as in the instrument used for this study[135], or raising the potential on the sample plate. When this occurs, ions with lower initial velocity and located closer to the sample plate will be affected by a greater

potential than the ions with a higher initial velocity, thus compacting the ion packet as they move into the field free region.

The delayed extraction parameters, set by ion source voltage 2 (IS2) and its corresponding pulsed ion extraction delay time, have a significant influence on the resolution and must be optimized for the specified mass range. A larger mass ion has greater energy and larger energy distribution. Therefore, the larger the ion mass is, the longer the delay time or the steeper the potential slope (IS2) needs to be to compensate for the large velocity spread and condense the ion packet. Both the voltage and the delay time are carefully tuned for optimizing the resolution in a given mass range. Parameters chosen for this study are defined in the "Materials and Methods" section.

Concepts of Experimental Design

After developing the core methodologies, a crucial point to consider is the experimental design. If not carefully planned, painstaking experiments with great potential could prove of little value due to potential sources of bias that can distort results. Much of this involves thorough application of three principles: randomization, replication, and blocking[136]. Therefore, when planning an experiment, several aspects of experimental design, as well as sample collection must be considered to achieve quality results that indicate differences due to true biological changes.

There is no common protocol by which repository facilities collect tissue samples; therefore, results from a study that utilized only one collection facility may not be the same as results obtained from a duplicate experiment using specimens collected at a different institution. Because of this limitation, when planning a study with large numbers

of samples, it is desirable to utilize multiple facilities. This minimizes institutional biases and reduces the possibility that biomarkers identified from a set of samples are limited to a particular institution. In this study ccRCC samples were collected from multiple facilities, though the majority of specimens came from two facilities.

The order in which samples are analyzed should be considered, because proteomic profiles may not be reproducible over time due to inherent instrumental variability, such as loss of laser performance, dirty sources, or loss of detector sensitivity over time. For this reason, it is ideal that investigators include some specimens in each experiment from each group being compared. In other words, in an experiment such as this, where tumor stage is a biological question of interest, it is ideal to run samples at random rather than analyzing all stage I's or all stage IV's, etcetera, at the same time or in the same order. Furthermore, analyzing all samples as they come in is not optimal randomization either. For studies in which all specimens are collected prior to the start of the research, it is advantageous to sample at random for the analysis, where no bias is given to a certain biological subset or collection facility. In studies such as this, when time constraints are present and samples are consistently coming in, samples should be randomized as much as possible. Randomization was difficult to achieve in this study; however, samples retrieved from storage for analysis were chosen independently of their diagnosis.

A standard operating procedure that is strictly followed will reduce the number of variables or biases that are of concern during the data analysis (discussed in Chapter III). The establishment of a single protocol to follow throughout the study is entirely under the control of the investigator. The experimental protocol for a study such as this involves:

sectioning the biopsy; washing the tissue; matrix application; mass spectral analysis, including calibration, which should be done prior to every experiment; pre-processing the data; and statistical analysis. The general standard operating procedure schematic for this experiment is shown in Figures 11 and 12.

In summary, experimental designs and implementations that are consistent, where protocols were identically implemented and strictly followed for every sample will yield results that indicate true biological effects. These experiments must include randomized analysis of specimens collected and quality control materials that achieve good calibration that is consistent for every analysis. Samples should be run multiple times to ensure the stability of a profile within a given specimen. Finally, following data collection, the spectra must be preprocessed identically, as discussed in the next chapter.

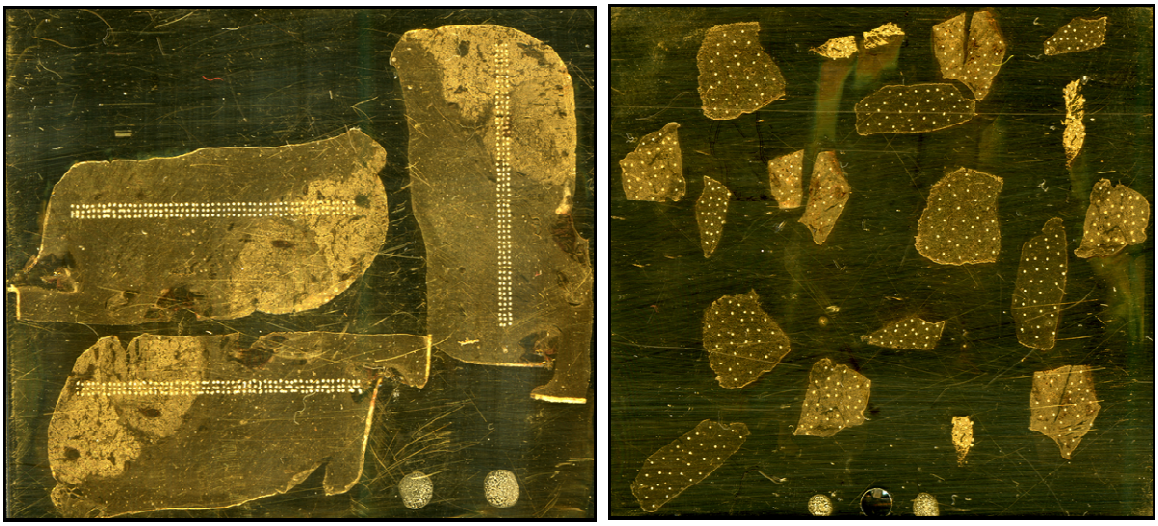


Figure 11: Experimental design and sample layout. Two types of specimens were collected for these studies: tumor and attached adjacent normal, patient-matched but separated tumor and normal. The first image (left) illustrates the layout of the tumor and attached adjacent normal samples, which were randomly placed, in triplicate, on the plate to avoid sampling bias when acquiring data in the mass spectrometer. Matrix spots were deposited in a thin array spanning the entire section, allowing the collection of data from tumor, tumor margin, and normal regions. The patient-matched but separated tumor and normal (right) were also randomized on the target plate. Typically three (as shown) or four patient's biopsies, both tumor and normal, were sectioned in triplicate and randomly arranged on the plate. Rather than in an array format, as shown on the left, the matrix spots were randomly scattered over the surface of each section.

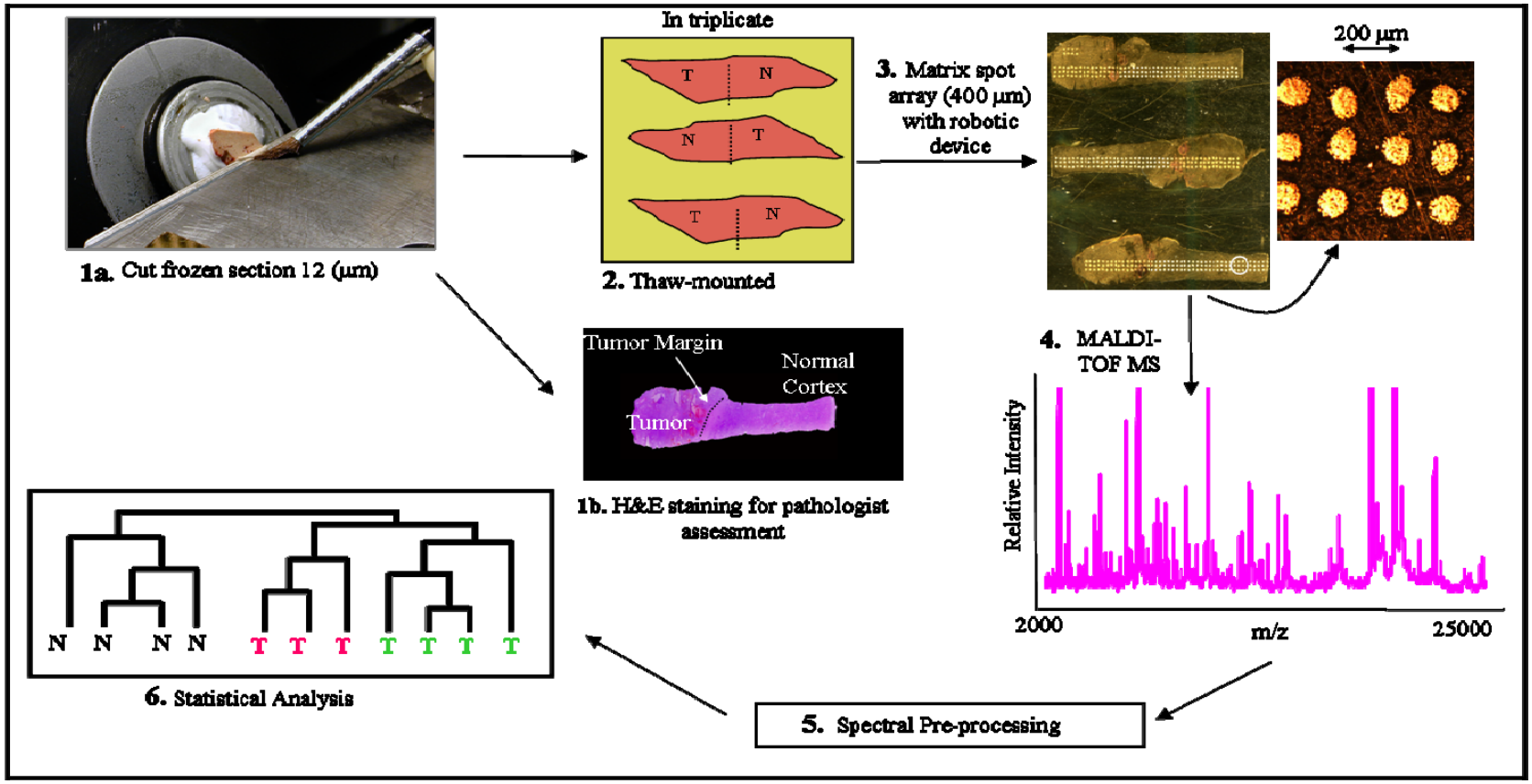


Figure 12: Experimental design, protocol schematic. 1.) 12 µm sections were obtained from tumor biopsies using a cryostat with subsequent sections cut and mounted onto glass microscope slides for histology assessment; 2.) sections were randomized on the MALDI target plate to avoid sampling bias in the mass spectrometer; 3.) sections on the target plate were washed (EtOH) prior to matrix spotting with the ARM device; 4.) mass spectra were automatically acquired; 5.) Biodesix software was used to pre-process the data set (baseline subtraction, normalization, alignment) as described in Chapter III; 6.) statistical analyses were then performed on the processed data set as described in Chapter III.

Summary and Conclusions

Protein profiling directly from tissue surfaces by mass spectrometry provides an abundance of molecular information in a high-throughput format. This allows for relatively simple analyses of the molecular contents and their spatial orientation within biological samples in any given state (e.g. disease, cancer, toxicity). In this chapter, however, it is noted that sample handling and preparation are important steps in assuring the maintenance of tissue quality and integrity. With proper application of these methods, it is possible to achieve high quality data in a high-throughput fashion. The methodology discussed in this chapter has been successfully applied toward the molecular analysis of tumor, tumor margin, and normal regions of ccRCC.

Materials and Methods

The MALDI matrix, sinapinic acid (SA), Fluka brand, was purchased from Sigma-Aldrich (St. Louis, MO). Arcos Organics brand trifluoroacetic acid (TFA) and 200-proof ethanol were ordered from Fisher Scientific (Suwanee, GA). Tissues were sectioned on a Leica cryostat model CM3050 S (Leica Microsystems, Germany). MALDI-TOF mass spectra were acquired using an Ultraflex II TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA)[135], employing smartbeam™ technology[137] and equipped with a solid state laser (~355 nm). Data were obtained in positive linear mode, with an accelerating voltage of 25 kV and optimized delayed extraction parameters (IS2 of 23.5 kV and a delay time of 150 ns). In a raster pattern with increments of 50 shots/spot, 400 laser shots were averaged for each matrix spot at a laser frequency of 200 Hz.

CHAPTER III

PROCESSING AND STATISTICAL ANALYSIS OF MASS SPECTROMETRY DATA

Profiling MALDI MS technology has contributed to the field of clinical proteomics by facilitating the discovery of disease-specific proteome signatures, which have the potential to aid in clinical diagnosis. The discovery of these signatures also contributes to a better understanding of disease mechanisms and provides prospective therapeutic targets. Although direct tissue analysis by MALDI MS technology requires a relatively simple sample preparation, its high-throughput platform produces large, complex data sets that require robust bioinformatics tools to reduce data complexity and mine mass spectral information to extract features (significant m/z values) that distinguish between the biological groups studied. Described in this chapter is the process involved in preparing spectra for statistical analysis followed by the statistical algorithms utilized to mine data for spectral signatures that differentiate the disease states compared.

Processing of Protein Mass Spectra

Spectral processing, also termed pre-processing because it is performed prior to statistics, is carried out on all data to reduce inter- and intra-experimental variance. Variations are produced from ionization effects, background noise, and calibration offsets, which result from biological or sample preparation effects. To reduce the effect of spectral variation on statistical analysis, mass spectra are prepared by removing background noise, normalizing intensities, and performing an additional calibration.

Correcting for these deviations enhances the ability of statistical algorithms to determine biological differences rather than experimental variations.

Baseline Subtraction and Noise Detection

Prior to spectral recalibration, normalization, and assignment of peak intensity values, all spectra must undergo background subtraction. Spectra acquired from MALDI MS profiling on tissue possess a severe chemical background, mostly in the low m/z region from salts and matrix cluster ions that must be quantified before it can be removed. In order to design the appropriate algorithm, the features of the background must be understood.

Mass spectra generated from biological tissue produces a baseline of variable intensity and shape that manifests itself in three general ways. As demonstrated by Norris and colleagues[138], the background can be a near-zero constant or a slight to intense monoisotopic decay, and in some cases, a bump is observed in the intermediate low mass range (e.g. m/z 3000-5000). As a result of the variable background shape from spectrum to spectrum, a one-value-fits-all strategy to removing it cannot be applied[138, 139]. Rather than designing fixed value, fit-model functions, the desirable baseline subtraction algorithm would be more flexible and easily adapted to specific features of each spectrum.

Despite several attempts to determine an algorithm for background subtraction, there remains no collective approach; however, there are three necessary criteria for these algorithms[140-142]. First, algorithms should avoid scenarios where splitting of spectra into multiple m/z sections is required to perform baseline subtraction. This approach

necessitates stitching spectra back together, which is adds complexity and time. Additionally, software packages should utilize spectra in an ASCII text format, so that spectra from any type of instrument can be easily exported to an ASCII format and baseline subtracted. Since MALDI-TOF MS offers a high-throughput platform for data collection, the processing software must also offer the capability.

Various commercially available software packages with baseline subtraction algorithms have been examined[138]. The commercially available software package that offers these capabilities, as well as a locally-flexible baseline subtraction algorithm is ProTS Data (Biodesix, Steamboat Springs, CO). Though the algorithm itself is proprietary, some generalizations can be made. The size of the sliding window for background subtraction is defined by user determined estimators (10 times the peak width at FWHM) at various m/z values across mass range allowing the program to determine the background for all subsets of m/z ranges. This gives the program more flexibility for different baseline trends [Figure 13]. It is probable that a best-fit regression or lowess-like line is estimated for each m/z neighborhood in each m/z range as defined by the user. Because peak width is one of the parameters used to estimate the baseline, the method will work on variable spectra acquired at different times assuming the spectra have similar resolution at any given m/z region. ProTS Data offers the ability to quickly compute and export the processed data in a batched process, and its required ASCII format provides for the input of a variety of spectra from different mass spectrometer manufacturers.

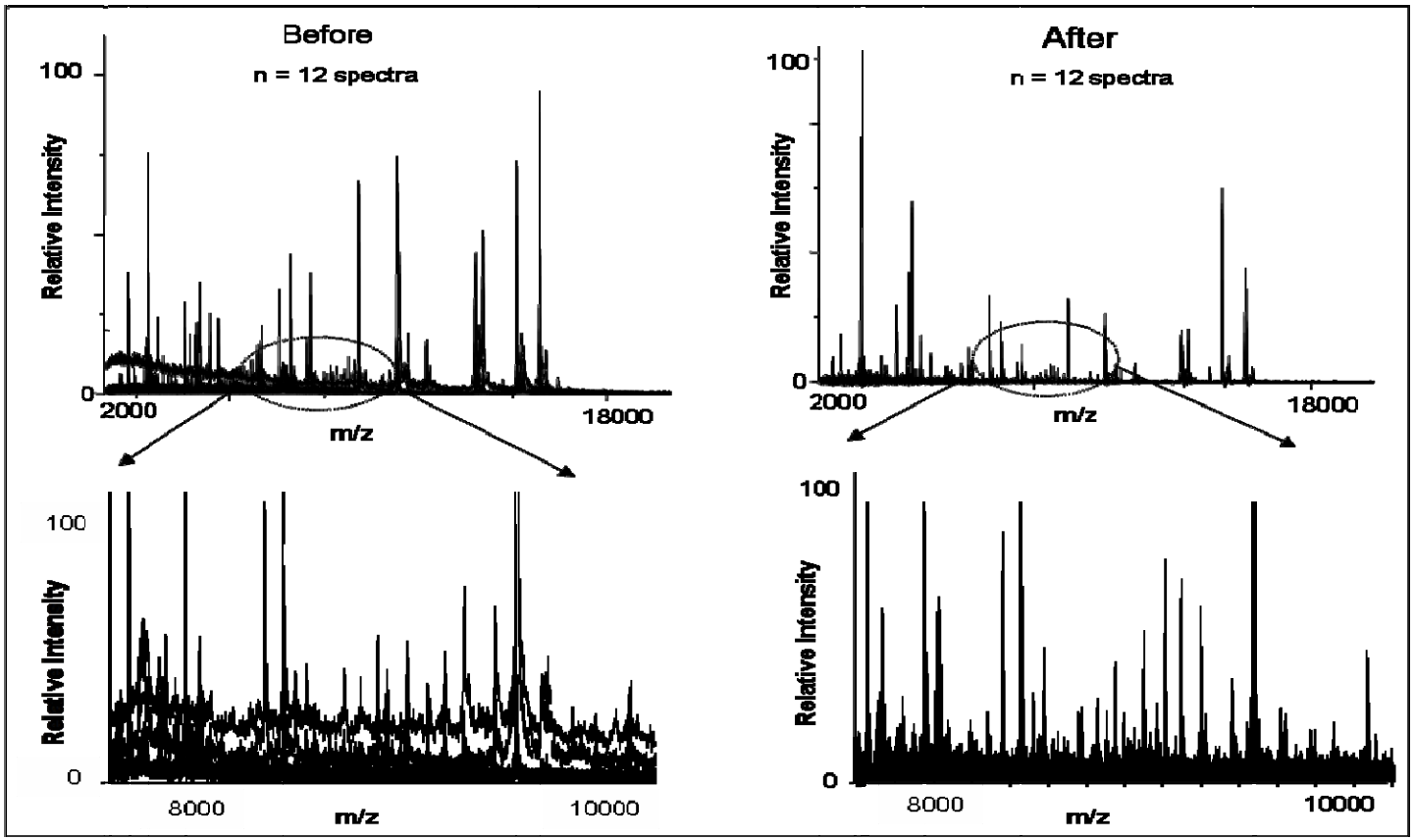


Figure 13: Background subtraction. The spectra above show an overlay of 12 spectra from 12 different patients. The left panel shows the inconsistent and high baseline prior to processing with background subtraction. The right panel shows a more uniform baseline resulting from processing with background subtraction.

Normalization of Intensity

After background subtraction, spectral normalization is needed to minimize spectrum-to-spectrum variations in peak intensity. Deviations arise from instrument changes that may include instrument retuning and laser replacement, uncontrollable variations in sample preparation, or biological variability. Eliminating variances will enhance the ability of statistical algorithms to extract feature patterns of biological significance from absolute peak intensities.

Several normalization methods exist, which fall into two categories. Spectra can be normalized to a defined reference spectrum, such as an average representative of the dataset, or to some reference that is independent of the collective data set, as in a constant-value total ion current (TIC). The methods are comparable, but the TIC method eliminates the need for creating an averaged representative from the data set. A comparison of multiple normalization methods has been performed[138], which included: cube root, log, and ln transformations; scaling to a set total ion current value; and scaling to the constant, determined noise. Norris and colleagues confirmed that the use of total ion current or noise constants for normalization produced the highest correlation and linearity of intensities. ProTS Data, the software used for baseline subtraction, provides a batch export platform for intensity normalization [Figure 14]. The user defines the total ion current normalization value, which is arbitrarily chosen at a fixed value greater than individual spectra.

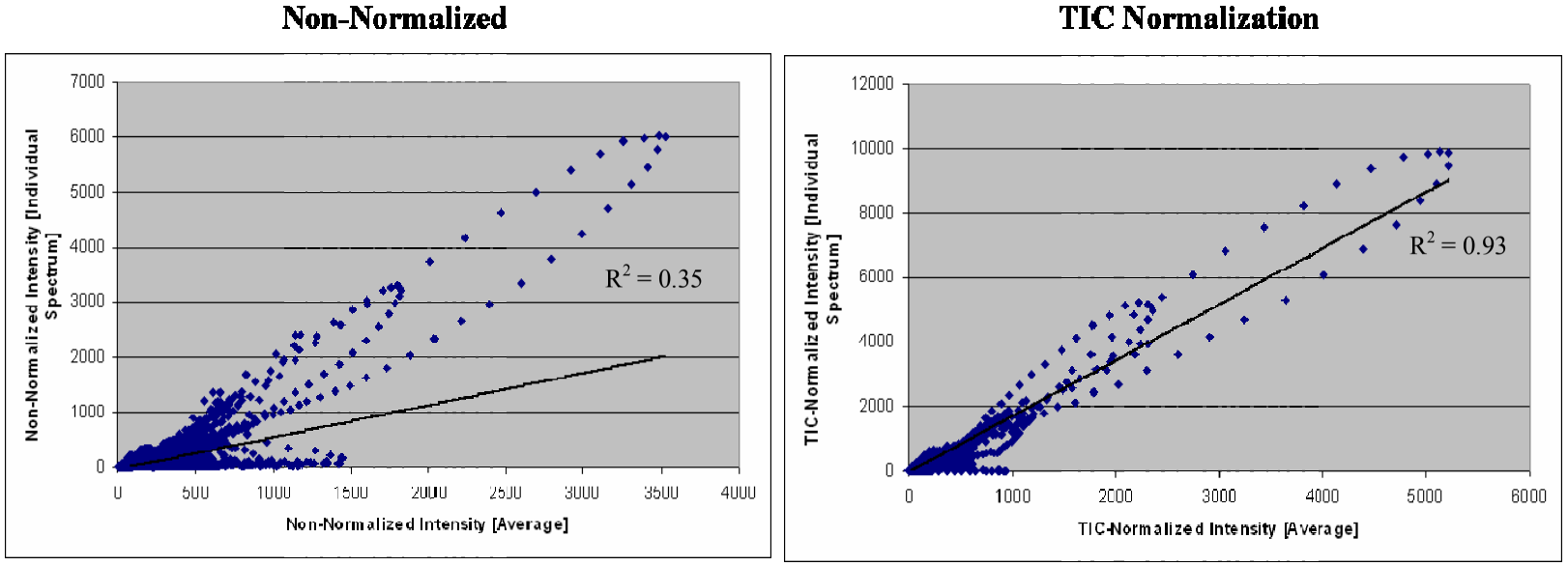


Figure 14: Effect of TIC normalization. In both plots, single spectra are compared against a group average. As shown, prior to normalization, intensity variations deviate significantly from linearity; however, after normalization, individual spectra and the group average have a more linear relationship.

Peak Detection

A prerequisite to spectral realignment is the separation of protein signals from background noise in the protein intensity measurements. For low resolution, high m/z spectra generated from biological profiles, the simplest method to identify peaks is to find the local intensity maxima that exceed a threshold value. Similar to its background subtraction algorithm, ProTS Data uses a locally adapted background algorithm to differentiate peaks from irrelevant noise. Through user-defined settings, which include peak width at m/z values across the spectrum and a signal-to-noise sensitivity threshold, the algorithm can correctly identify excessive noise between large peaks, while retaining sensitivity for less intense peaks.

Smoothing the spectra prior to the processing steps will further reduce noise and enhance the ability of the algorithm to define peaks. By further minimizing electrical and chemical noise, smoothing will improve peak signal-to-noise ratios, enhancing the ability of peak detection algorithms to correctly identify and label the local peak maxima. For this study, smoothing was performed during acquisition, with a low-pass bandwidth filter on the digitizer, which discards high frequency noise and improves signal-to-noise in the m/z range collected without sacrificing resolution [Figure 15]. When low m/z peptide spectra are of interest, smoothing should be performed cautiously, as it may degrade isotopic resolution. This study focused on a higher mass region, from 2000-30000, which contains protein/peptide signals of low resolution. Inclusion of a low-pass filter ultimately enhanced the detection of peaks low in intensity.

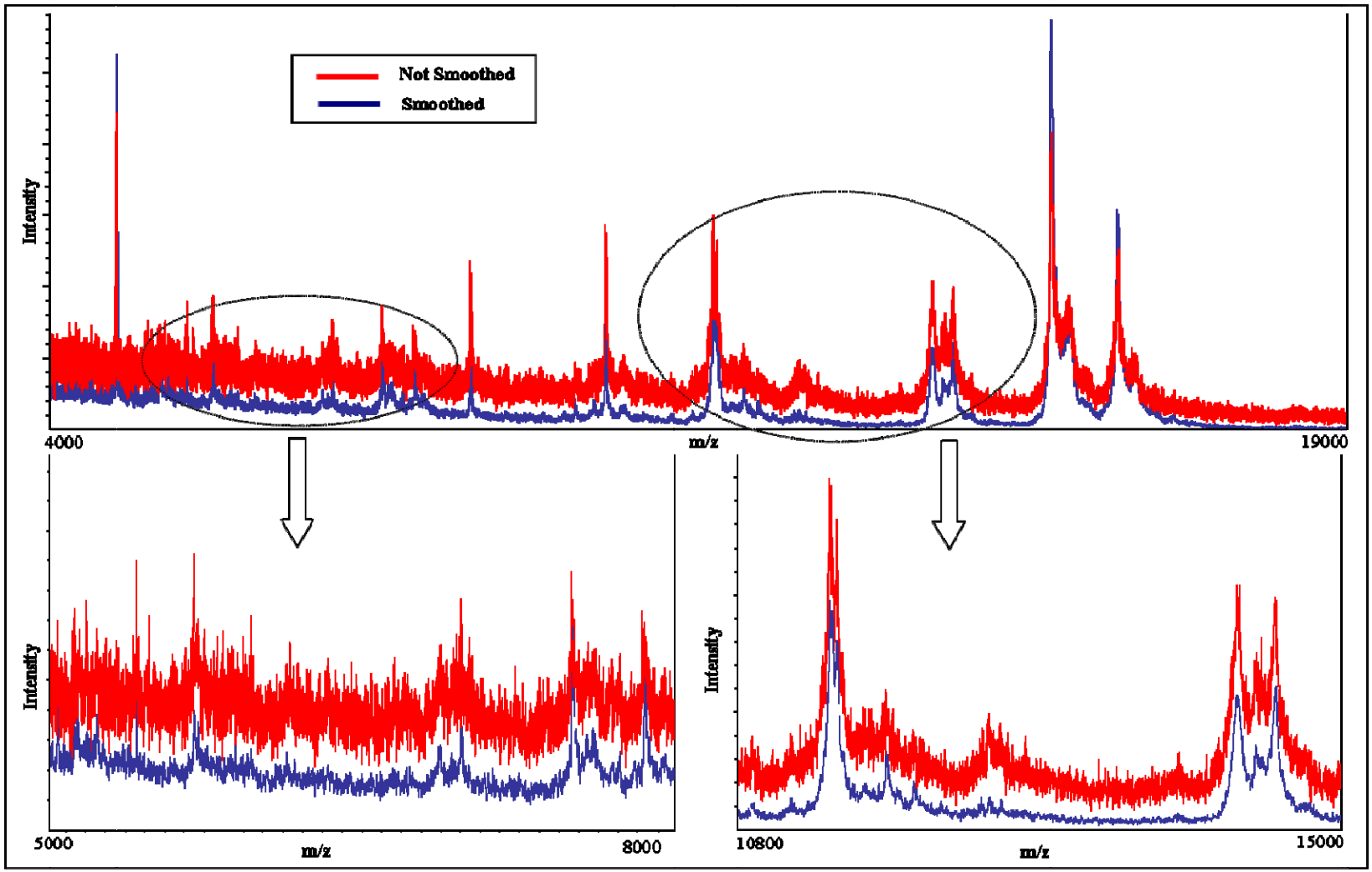


Figure 15: Noise filtering. This figure illustrates the effect of the filtering option on the mass spectrum. As shown, the smoothing parameter improves the S/N and enhances low abundant signals in the mass range analyzed.

Spectral Calibration/Realignment

High-throughput MALDI-TOF MS experiments, where multiple samples are analyzed, on different days, require calibration for precise alignment of experimental data in order to perform future statistical analysis and extract protein signatures that differentiate the groups compared. In each experiment, samples are placed on one of multiple, but theoretically similar MALDI plates, which are then placed onto a moving stage in the mass spectrometer. Depending on the sample position, the alignment of spectra may shift on different regions of the plate and between samples on different plates. This deviation in calibration may be due to several factors, which include irregularities in plate flatness, sample topographies and thickness, and matrix crystallization. External calibration alone may produce m/z errors up to 500 ppm, depending on the MALDI-TOF instrument. Although standard calibrants are useful for an initial alignment of spectra, additional methods are needed to reduce the error, which may cause problems for qualitative comparisons.

Calibration of MS signals can be performed by external or internal calibration. Both methods rely on known m/z values of standard peptides/proteins to calibrate spectra to common coordinates. External calibration is performed by placing a mixture of known calibrants, spanning the mass range of interest, on the sample plate away from the analyte. Although external mass calibration is performed prior to each experiment, inter- and intra-experimental mass variations still exist, thus, it is necessary utilize an additional approach. Internal calibration, where calibrants are mixed with the analyte prior to analysis, seems advantageous, but signals from the calibrants are often suppressed by the analyte ions and vice versa. There is also potential for calibration signals to overlap with

the analyte signals, resulting in erroneous m/z assignments. A post-analysis approach to internal calibration is possible using known peaks in the analyte. Unfortunately, when dealing with biological samples, it is unlikely that a sufficient number of peaks can be accurately assigned to their known m/z values, given that most of the peaks have not been identified in that particular sample. Ultimately, an external calibration is essential for an initial calibration and an assessment of instrument performance, however, it is necessary to utilize an additional approach in conjunction with this method.

A practical alternative approach, in addition to external calibrants, would be to determine a subset of peaks in common to the dataset, and use these as the reference peaks for realigning the spectra. Several algorithms have been established, but few are commercially available. In this study, ProTS Data was utilized for spectral realignment. Though its algorithm is proprietary, some generalizations can be made on how the program is designed for realigning spectra. In brief, representative spectra are loaded into the program, which mines the data to determine peaks in common (also termed peak clusters) based on user-defined thresholds. The thresholds define the m/z error window (e.g. +/- 200 ppm) in which to search for peaks of a given m/z. Depending on the number of spectra generated in a study, only a portion of the spectra may be used to define the peak clusters. In this case, a pooled set of spectra from each sample can be used to represent the dataset. A list of the peak clusters and the corresponding number of spectra containing them is presented to the user, who manually determines what peaks to use as recalibration features. Chosen peaks must span the entire mass range investigated to prevent erroneous assignment of m/z values outside of the chosen calibration points.

Once the calibrants are defined, the calibration peaks are assigned an arbitrary mass, defined as either the median or mean value of that peak for the pooled data-set representative, and each spectrum is realigned using a quadratic-fitting calibration algorithm. An example is of spectra before and after alignment is shown in Figure 16. The user may set additional tolerance thresholds for detecting peaks in the spectra that should be considered as internal calibrants. This approach can reduce m/z shifts by 5- to 10-fold. Depending on the instrument and its corresponding calibration error, the realignment can produce inaccurate m/z assignments; therefore, confirmation of m/z assignments from the raw data is needed for peaks determined significant from the statistics. In this dataset, it was observed that the raw data was well calibrated and required minimal realigning, thus peak assignments before and after realignment were no more than an average of 0.5 to 2 mass units different.

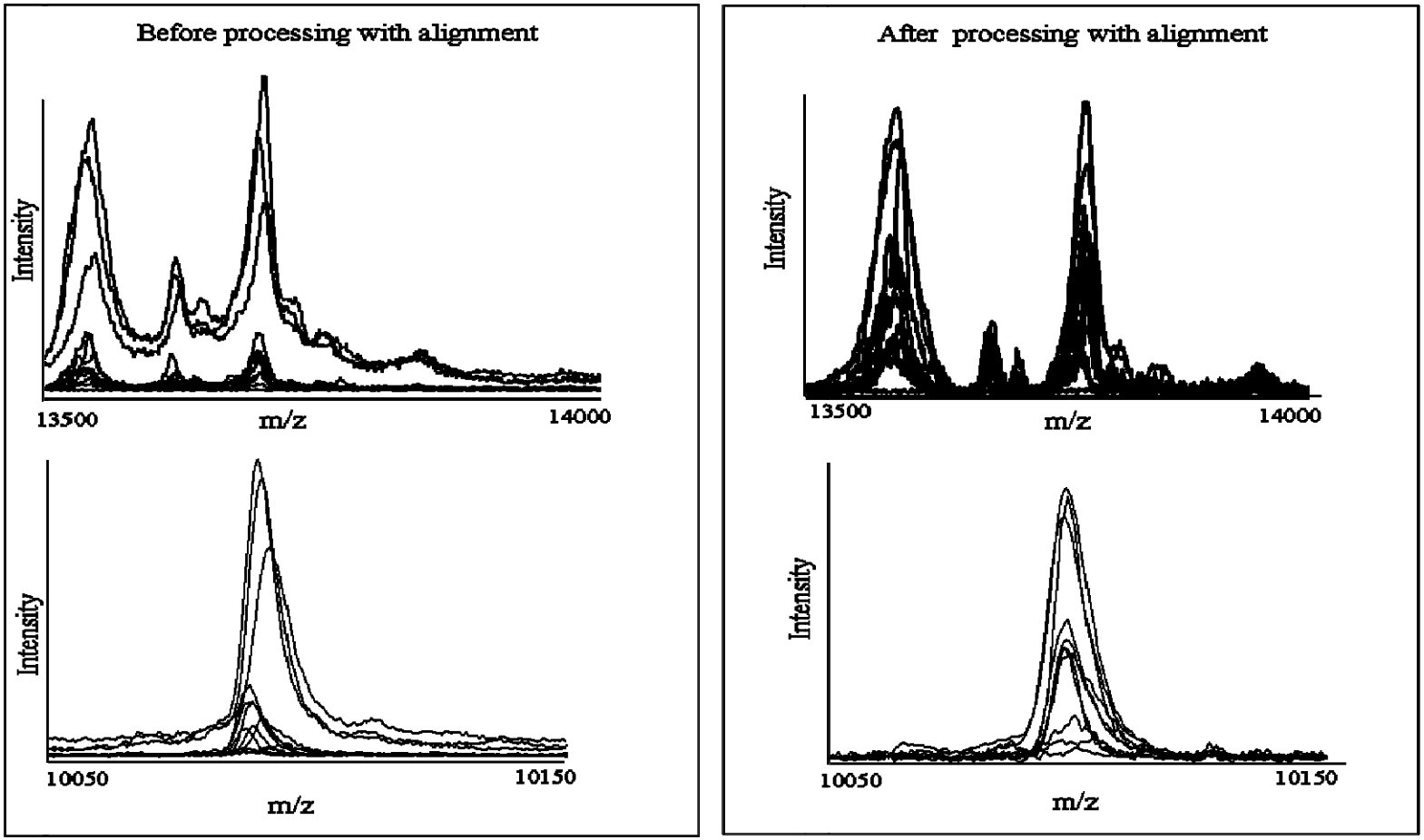


Figure 16: Processing with realignment. The spectra above show an overlay of 12 spectra from 12 different patients. The left panel shows the peak alignment prior to processing with the re-calibration feature. As illustrated in the left panel, multiple m/z ranges may have different initial calibration errors, necessitating a flexible calibration algorithm. The panel on the right illustrates a more uniform alignment after processing with the re-calibration algorithm.

Peak Binning

Although mass spectra have been recalibrated, sometimes m/z variations remain that must be minimized prior to statistical analysis. Once peak detection parameters are defined, processed spectrum files are reduced from a continuous mass spectrum of ~32,000 data points and exported as ASCII peak list (~300 peaks) files containing the peak m/z value and its corresponding intensity. Peaks were aligned or grouped by a series of peak bins or mass windows that varied in size, increasing proportionally with m/z values. Once the peaks were binned, a spreadsheet of the results was created. The first columns of the matrix sheet specified the beginning and ending mass of each bin. Each row represented one file or patient; the data in each patient column listed the peak value in each of the bins. A separate matrix file was created that included peak intensity information. These files were utilized in preparation for statistical analysis.

Statistical Analysis of Processed Data

After spectra are processed, statistical analyses are performed on MALDI MS profiling data to discover spectral features that that distinguish between biological states. Often, the goal of statistics is to extract individual m/z features or abundances that differentiate between samples that have been grouped according to their pathological classifications. In the case of cancer, for example, information obtained is used to develop a model that will aid in predicting a tumor's molecular classification or a patient's chance of survival based on the MS protein pattern of a biopsy.

The goal of this research was to develop and apply an approach by which to examine the molecular characteristics of tumor margins in a particular cancer system,

clear cell renal cell carcinoma. The statistical analysis of protein mass spectra from this data set concentrated on two aims: 1.) determine if there are molecular features that differentiate between pathological classifications, and 2.) determine if the molecular environment outside of the histological tumor margin expresses tumor characteristics. Data is generally separated into two groups, a training set (supervised analysis) to develop the classification model and a test set (unsupervised analysis) to verify the prediction accuracy of the model. The statistical validation approaches will also be discussed.

Though numerous and often complex, statistical approaches have been applied to the analysis of complex, two-dimensional data, such as microarray and mass spectrometry data. This section describes the independent algorithms used to analyze the data set produced in this study: significance analysis of microarrays (SAM) and the permutation *t*-test. In any statistical platform, it is advantageous to perform more than one statistic test. In this study, results from each statistical method were used to validate and compare lists of significant features. Both SAM and permutation *t*-test algorithms can be applied to pairwise comparisons, which minimizes the influence of biological variations.

Significance Analysis of Microarrays

Significance Analysis of Microarrays is a statistical technique for finding significant genes in a set of microarray experiments. Originally designed for microarray data, this statistical method can easily be adapted to protein expression data; both result in a measure of expression level for any given species (gene or protein). Detailed information on the SAM method can be found elsewhere[143], but is summarized here.

SAM assigns a score of relative difference $d(i)$ to each species (i) based on the ratio of change in species (m/z value) expression to standard deviation in data for that species:

$$d(i) = \frac{\bar{x}_1(i) - \bar{x}_2(i)}{s(i) + s_0}$$

where $\bar{x}_1(i)$ and $\bar{x}_2(i)$ are defined as the average levels of expression for species (i) in class 1 and class 2, respectively (e.g. tumor and normal; stage II and stage IV). The standard error of difference of mean levels of expression for species (i) is represented by $s(i)$ and estimated by the following equation:

$$s(i) = \sqrt{a\left\{\sum_m \left[x_m(i) - \bar{x}_1(i)\right]^2 + \sum_n \left[x_n(i) - \bar{x}_2(i)\right]^2\right\}}$$

where $\sum_m$ and $\sum_n$ are summations of the expression measurements for samples in class 1 and class 2, respectively, $a = (1/n_1 + 1/n_2)/(n_1 + n_2 - 2)$, and n_1 and n_2 are the numbers of measurements in class 1 and class 2, respectively. At low expression level, the variance of $d(i)$ can be high due to small $s(i)$. A positive constant, s_0, is chosen to minimize the coefficient of variation of $d(i)$, hence making the variance of $d(i)$ independent of species expression level.

Species are ranked by the magnitude of their $d(i)$ values, so that $d(i)$ is the *i*th largest relative difference. In order to assign statistical significance to the ordered relative difference, $d(i)$, SAM computes the expected relative difference $d_E(i)$ under the null hypothesis that there is no difference in species expression level between class 1 and class 2. SAM permutes[144] the data by scrambling the label of samples and hence assumes that all null hypotheses are true and there are no differences in the expression levels of all

species between class 1 and class 2. For each of such permutations, SAM computes a score of relative difference $d_P(i)$ for each species and ranks the scores by their signed magnitude. The expected relative difference $d_E(i)$ under the null hypotheses is defined as the average of the ranked scores of relative difference over all permuted data.

$$d_E(i) = \sum_p d_p(i) \,/\, (\#\text{ of permutations } p)$$

The species with relative difference $d(i)$ deviating from its expected value $d_E(i)$ under the null hypotheses by a distance greater than a threshold Δ is called significant. As Δ decreases, the number of species called significant will increase at the expense of increasing the number of potential false positives.

SAM uses a false discovery rate (FDR) to correct for multiple comparisons. The false discovery rate measures the proportion of false positives among all species called significant. For a fixed threshold Δ, SAM estimates the false positives by computing the average number of species called significant over all permuted data, and then divides it by the number of species called significant from the data with original sample labeling. SAM estimates FDR for a sequence of different thresholds (Δ), and the optimal threshold (Δ) will be the one with estimated FDR equal to or smaller than a pre-specified acceptable FDR.

For paired data, the expression level of each species (i) is associated with an index vector $\{-1,1,-2,2...-K,K\}$. Observation $-K$ in class 1 is paired with observation K in class 2. SAM assigns a score of relative difference $d(i)$ to each species (i) defined by the equations:

$$z_{ik} = x_i(k) - x_i(-k)$$

$$r_i = \frac{\sum_k z_{ik}}{K}$$

$$d(i) = \frac{r_i}{s(i) + s_0}$$

where $x_i(k)$ and $x_i(-k)$ are defined as expression levels of species (i) for paired observations $-k$ in class 1 and k in class 2, respectively. K is the number of paired samples. The standard error of mean difference in expression levels between paired observations $-k$ in class 1 and k in class 2 of species (i) is represented by $s(i)$ and determined by the following equation:

$$s(i) = \sqrt{\frac{\sum_K (z_{ik} - r_i)^2}{K(K-1)}}$$

The positive constant, s_0, is chosen to minimize the coefficient of variation of $d(i)$ as described above. For paired data, permutation is performed by randomly scrambling labels of observations within each $-k$, k pair.

Permutation t-test

Detailed descriptions of the permutation *t*-test may be found elsewhere[145], but is describe here in brief. The permutation *t*-test utilizes the standard two-sample *t*-test, where the *t* statistic is defined as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\bar{x}_1$ and $\bar{x}_2$ are mean expression levels of species in class 1 and class 2, respectively, and n_1 and n_2 are number of samples in class 1 and 2, respectively. The pooled estimate of standard deviation, s_p , is defined as:

$$s_p = \sqrt{\frac{(n_1 - 1)(s_1^2) + (n_2 - 1)(s_2^2)}{(n_1 + n_2 - 2)}}$$

where s_1 and s_2 are estimate of standard deviation of expression levels of species in class 1 and class 2, respectively.

In the permutation *t*-test, the standard *t* statistic is computed on the log-expression of each species, because the *t*-test assumes that the data is normally distributed. Next, labels are randomly permuted among the samples and the *t* statistic for each species in the permuted data set is computed. This process is repeated 10,000 times. Finally, the *p*-value for each species is determined by computing the proportion of the *t* statistic for that species over all permuted data sets which have absolute values equal to or greater than the absolute value of *t* statistic from the originally labeled data set.

The paired data analysis using the permutation *t*-test is similar to that of SAM. For paired data, the expression level of each species (i) is associated with an index vector $\{-1,1,-2,2...-K,K\}$. Observation $-K$ in class 1 is paired with observation K in class 2. The *t* statistic for paired data is defined as:

$$z_{ik} = x_i(k) - x_i(-k)$$

$$r_i = \frac{\sum_k z_{ik}}{K}$$

$$t = \frac{r_i}{\sqrt{\frac{\sum_K (z_{ik} - r_i)^2}{K(K-1)}}}$$

where $x_i(k)$ and $x_i(-k)$ are defined as expression levels of species (i) for paired observations $-k$ in class 1 and k in class 2, respectively. K is the number of paired samples. For paired data, the permutations are performed by randomly scrambling labels of observations within each $-k$, k pair.

The *p*-values from permutation *t* test are further adjusted to correct for multiple comparison using the method proposed by John Storey[146]. This method estimates *q*-value for each species from its corresponding *p*-value, which is obtained from a certain test. The *q*-value for a particular species is the minimum false discovery rate that can be attained when calling all species with equal or smaller *q*-value significant.

Statistical Validation and Classification Prediction Accuracy

The final step of a statistical analysis is validating the statistical results, which can be performed by two different methods. The conventional and ideal method is to split the dataset into two groups: 1.) a training set, in which the data are arranged into groups based on their classifier (e.g. cancer stage/phenotype); 2.) a test set, in which the data are not arranged into groups. The latter set is used to determine how well the statistical algorithms designed around the training set will classify the data from the test set into the correct classifier groups.

In this study, the bootstrap procedure[147] was used to validate the process of feature identification. In cases where the sample size is under a couple hundred samples,

the bootstrap method is superior to the aforementioned data splitting method, and instead, utilizes the entire data set. For this work, the original data were randomly sampled with replacement to simulate a virtual sample of size *n*, where *n* is equal to the size of the original data set. For each of these bootstrap samples, the same statistical analysis procedure was followed to identify the significant feature, thus the SAM and permutation *t*-test statistics were employed. Features identified as significant by both methods with a false discovery rate (FDR) less than 0.01 were considered as potentially significant features. This process was repeated for 100 bootstrap replications in order to observe how the features behave over the 100 repetitions. A highly consistent list of significant features from the 100 bootstrap replications implies good reproducibility of the result from the original data and highly consistent performance of the statistical analysis procedure.

After validating each feature, the leave-one-out cross validation (LOOCV) method[147] was used to assist in evaluating the predictive power of each significant feature. For each significant feature, one sample was left aside and a classifier, using this feature as input feature, was trained on the other *n*-1 samples. The trained classifier was then used to predict the class of the left-out sample. This process was repeated *n* times to obtain an estimate of the prediction accuracy of the feature. The support vector machine (SVM) was used as the classifier. SVM projects the training data from the input space to a new feature space with higher dimension and tries to find a hyperplane separating the training samples in the feature space and establish a class prediction rule. Once the rule is determined, it can be used to predict the class of new sample with unknown class; thus, the classification prediction accuracy of that feature is between 1 and 100 percent. Once

all of the features have been determined, this same procedure is used to determine the class prediction accuracy of a group of features, which will be more powerful for separating two groups than a single feature.

Once the features used to classify two groups (e.g. tumor and normal) were determined multidimensional scaling (MDS) was used to visualize the distances among samples. The *p* significant features were selected, and the pairwise Euclidean distances between *n* samples $x_{1,}x_2,...,x_n \in \mathbf{R}^p$ were computed:

$$d_{ij} = \left\| x_i - x_j \right\|$$

d_{ij} :distance between samples *i* and *j*

Multidimensional scaling projects data of *p* dimensions into a lower *k*-dimensional ($k < p$) space while trying to preserve the entire original pairwise distance structure. For example in classical metric scaling, MDS seeks values $z_{1,}z_2,...,z_n \in \mathbf{R}^k$ ($k < p$) to minimize a stress function:

$$\sum_{i \neq i'} \left(s_{ii'} - \left\langle z_i - \bar{z}_i, z_{i'} - \bar{z}_i \right\rangle \right)^2$$

where $s_{ii'} = \left\langle x_i - \bar{x}_i, x_{i'} - \bar{x}_i \right\rangle$, the centered inner product of x_i and x_j

The first two MDS coordinates from classical scaling can be plotted to show the separation of clusters of samples.

Alternative Statistical Approach

In addition to the statistical approaches mentioned above, there was a second source of computation. The company, Biodesix, Inc., was also utilized for an independent assessment of the data in an effort to obtain further confirmation of the results. They used

the same preprocessing software, with slightly different settings. They did not, however, use the same statistical approach. Their services were not used for every biostatistical computation, but when appropriate the results of their analysis can be found in the Appendices. In brief, their methods are described below.

Raw spectra were pre-processed using ProTS Marker (Biodesix Inc., Steamboat Springs, CO), a modified version of ProTS Data. Significant features were identified using Wilcoxon rank sum test and sorted by their *p*-values. Features having *p*-values less than 0.0005 were considered candidate features to be included in the classifier. ProTS Marker uses a modified version of *k* nearest neighbors (*k*NN) as classifier. Given a query sample, *k*NN finds the *k* samples in the training set closest in distance to the query sample, and assigns a class to the query sample using the majority vote among the *k* neighbors. In case there is a tie, *k*NN will assign an undetermined class to the query sample. The list of significant features to be included in classifier and the optimal value of *k* were chosen to minimize the leave-one-out cross validation (LOOCV) test error.

Analysis Workflow

An overview of the workflow is illustrated in Figure 17. Raw mass spectra were converted to ASCII text files prior to any processing. One spectrum was then opened into ProTS Data (Biodesix Inc., Steamboat Springs, CO), the processing software, to estimate each of the parameters for noise detection, peak detection, and background subtraction. After an initial estimation, other individual spectra, from different regions or different patients, were imported to confirm or fine-tune the parameters for the configuration file. Approximately 100 spectra, representing each of the patient samples, were then used to

find peaks in common between spectra for the re-alignment. This file was linked to the configuration file, which contained parameters for background subtraction, noise estimation, peak detection, normalization, and reference peak values for re-alignment. In batch mode, all ASCII-formatted raw spectra underwent background subtraction, normalization, and re-alignment. Processed data was exported as ASCII format spectra files, which contained the entire spectrum of approximately 32000 data points.

Additional computations were performed in the R program, a freeware language platform for statistical computing and graphics[148]. Processed data was arranged into several groups for future comparisons: Tumor; Normal, Near Margin Tumor; Near Margin Normal; stages and grades I, II, III, and IV. Prior to statistical analysis, spectra were averaged, in the R program, from each patient to obtain an averaged spectrum per patient in each of the above categories. The averaged spectra were batch processed, using ProTS Data, to export an ASCII format peak list file, which contained the peaks detected by the peak detection algorithm, and their corresponding absolute intensity values. Peak list files were then binned using custom software, which exports a matrix file containing each sample and its peak list value in each bin; there was one matrix file created for each set of comparisons (e.g. tumor and normal; stage I, III, III, and IV). In the R console, using both the binned peak list matrix file and the spectrum files, the area under the curve of each peak in the list was calculated. A second matrix file was created, containing each peak and its corresponding AUC for every patient. The statistical analyses were then carried out in the R console using the AUC matrix file.

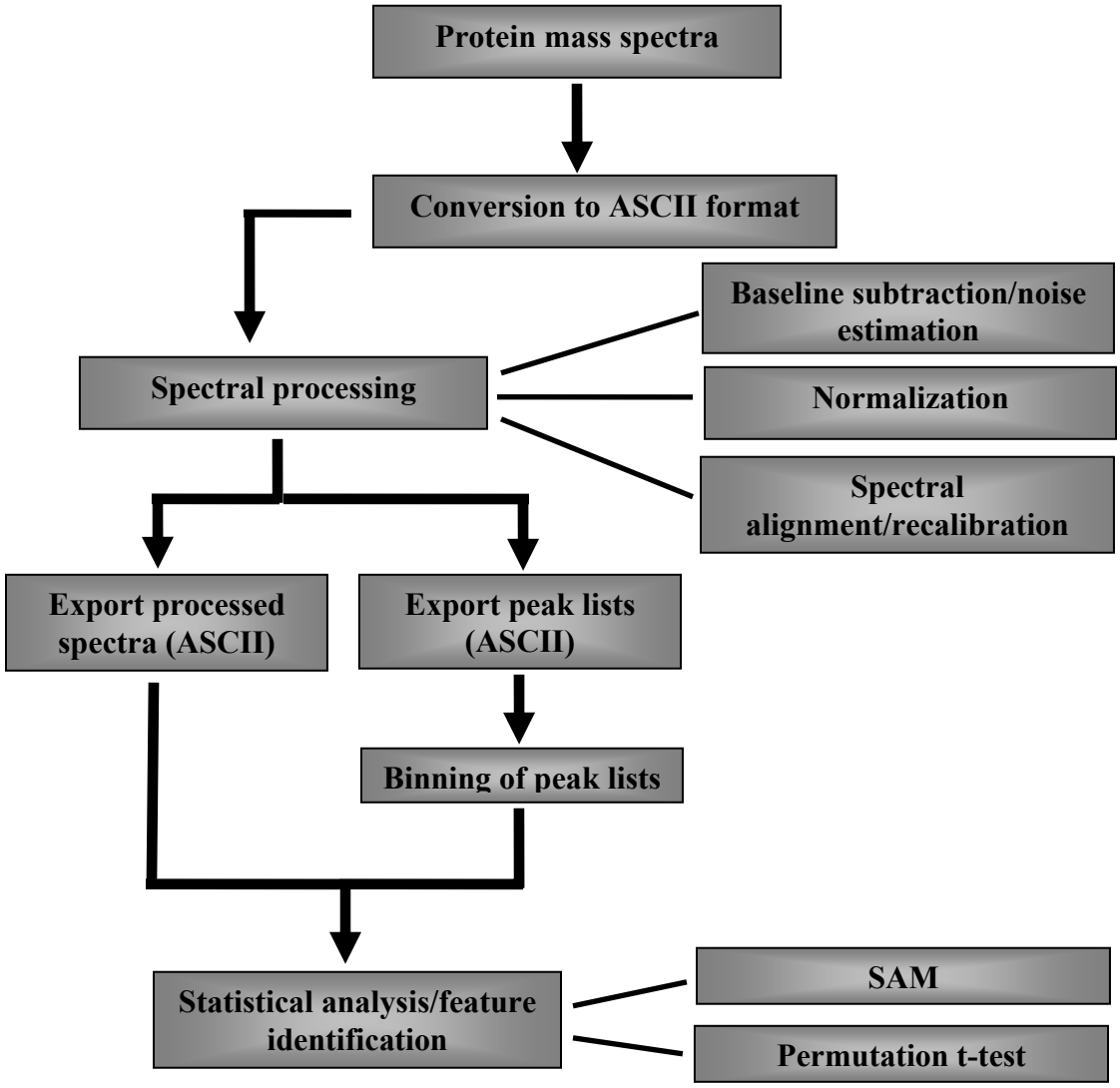


Figure 17: Data analysis workflow.

Summary

This chapter described the data-handling process involved in mining complex, proteomic mass spectra to extract biologically meaningful information. Spectra should be preprocessed to reduce variability and improve spectral quality. The same processing method should be applied to every spectrum, or additional variability will be introduced. Ultimately, when performed carefully, preprocessing the data will enhance the ability of statistical algorithms to identify spectral features of potential biological significance. The statistics described in this chapter were used to address several biological questions. The results of these analyses are described in Chapter IV.

CHAPTER IV

ASSESSING THE MOLECULAR CHARACTERISTICS OF TUMOR MARGINS BY MALDI MASS SPECTROMETRY

Ensuring successful tumor resection is a major concern in clinical oncology and pathology. Obtaining tumor-free surgical margins at the time of surgery is essential in decreasing tumor recurrence rates[106]. Currently, histopathological assessment of hematoxylin and eosin stained sections of the resected tumor is the primary method for assessing surgical margin status. Depending on the type of tumor, the margins analyzed can be taken 1 to 2 cm away from the gross tumor margin. If microscopic tumor is present at the surgical margins, the rate of local recurrence increases and the survival rate decreases[107], indicating current histological analyses may be inadequate in determining uncompromised margins. To circumvent this problem, more studies are needed to investigate the molecular characteristics that comprise tumors and the normal environments adjacent to them.

The idea of a "field effect" or "field cancerization" was proposed many decades ago [149], yet the molecular abnormalities that characterize this field are mostly unknown. Depending on the type of tissue, recurrence can occur a few months to a few years after removal of the primary tumor, yet there are several possible explanations, including: 1.) the resection of a histologically tumor-negative surgical margin leaves behind cells that have already undergone malignant transformation at the molecular level prior to resection; 2.) the presence of tumor-secreted proteins into the normal environment leaves behind an abnormal environment outside of the tumor-negative resection margin; 3.) the

presence of the tumor influences aberrant secretions in seemingly normal cells outside of the tumor margin. Without understanding how histologically normal cells become transformed into malignant cells outside the tumor margin and with the inability of pathologists to visualize this transformation with current histological procedures, molecular markers have the potential to enhance the way pathologists analyze margin status by providing an additional, and yet complimentary, way to identify aberrant tissue environments outside of the histological tumor margin[107].

Previous studies of tumor margins have focused on selective antibodies or single proteins, with little or no mention of how far the abnormality extends from the conventionally-defined tumor border[150-151]. High-throughput, global proteome discovery tools offer the potential to examine how proteins change in the environments surrounding the histologically-defined tumor boundary. To accomplish this task, proteomic methods must first be developed and applied to a model system to determine the effectiveness of the methodology and applicability to any system.

In this study, we have developed an *in-situ*, direct tissue profiling approach to test the hypothesis that there are aberrant molecular characteristics beyond the histologically-defined tumor margin. Matrix-assisted laser desorption ionization mass spectrometry is a high-throughput, sensitive tool that has been successfully applied to numerous clinical proteomics investigations and provided a plethora of potential biomarkers for numerous diseases from both *in-situ* and *ex vivo* conditions. Using this technology, this research focused on the analysis of tumor margins in clear cell renal cell carcinoma samples

Current surgical management of clear cell renal cell carcinoma is complete kidney removal, yet the surgical margin status is of increasing concern due to the rising demand

for nephron sparing surgeries. Nephron sparing surgeries remain rare, however, due to concern over local recurrence. Since the radical nephrectomy remains to be the dominant surgical procedure, ccRCC is an ideal system for this study due to the ease of acquiring the samples needed in contrast to other systems.

In this study, MALDI-TOF MS was used to examine molecular signatures in the histologically defined tumor, normal, and margin tissue environments of ccRCC samples. Overall, there were 75 samples analyzed from 75 different patients, each with matching tumor and normal tissues. This study determined that there were aberrantly expressed molecular characteristics outside of the histologically-defined tumor margin. Using specific features, expression patterns were mapped from the tumor into the normal tissue to determine how these patterns change with tumor aggressiveness, as defined by conventional pathological diagnosis. Furthermore, specific proteomic features were discovered that differentiate normal and tumor tissue, as well as between tumor stages and grades.

Results

Experimental Design and Sample Preparation

Seventy five tumor and matched normal tissues were collected from 75 patients. Thirty four of the 75 samples contained tumor with attached, adjacent normal tissue for molecular assessment of the histology-marked tumor borders. Each sample was sectioned for MALDI-TOF MS analysis. Serial sections were placed onto a microscope slide and stained with hematoxylin and eosin dyes for histology analysis. For each sample, a

pathologist marked regions pertaining to normal and tumor, and where applicable, the histological tumor border was marked. Regions of hemorrhage, necrosis, and visibly abnormal tissue due to high inflammation were marked. Tissues appearing to be poorly preserved were excluded. The sections for MALDI-TOF analysis were placed onto gold-coated MALDI plates and discrete matrix spots were placed in specific regions of the tissue to be analyzed [Figure 11]. *In situ* MALDI-TOF MS analysis of tumor, normal, and regions around the histological tumor border was performed. Overall, approximately 24,000 individual mass spectra were acquired. For statistical analyses multiple spectra from each patient and each tissue region were averaged to generate one representative peak list and corresponding intensity value per tissue region per patient. A general schematic of the experimental design is shown in Figure 12. For most matrix spots, approximately 200-300 individual m/z features were observed between the m/z range 2,000 and 25,000. To assess the variability of the data, the concordance correlation coefficient was determined for the tumor and normal tissue for each patient. For tumor tissue, the value ranged from 0.34 (poor) to 0.89 (excellent), with an average of 0.67 and a median value of 0.67, without excluding outliers. The normal tissue values ranged from 0.21 (poor) to 0.91 (excellent) with an average of 0.62 and a median value of 0.64, without excluding outliers. The inter-patient concordance coefficient was 0.6 for tumor and 0.6 for normal tissues. This indicates an acceptable level of inter- and intra-variability.

Molecular Analysis of Tumor Margins by MALDI MS

The overall objective of this study was to determine if there are aberrant molecular features present outside of the histological border and then to understand how the expression patterns of these features present themselves from tumor into the normal tissue. To accomplish this, the molecular profiles from the different regions of tissue were compared. For this assessment, only tissue samples consisting of tumor and attached adjacent normal were considered (34 samples) [Table 1]. Figure 18 highlights the regions of interest for the analysis. These four regions consist of far tumor, near margin tumor, near margin normal, and far normal. With the focus being on the molecular characteristics of the region outside of the tumor margin, tissue regions were grouped independently of their respective tumor grade or stage diagnosis.

To perform statistical analyses, the mass spectrometric data was processed, averaged by region for each patient, and grouped into four categories. A paired analysis was performed on the regions as follows: far tumor versus far normal (Figure 19, Table 2), near margin tumor versus far tumor, near margin tumor versus near margin normal (Figure 20, Table 3), near margin normal versus far normal (Figure 21, Table 4). MDS plots of the data sets compared in far tumor versus far normal and the far normal versus near margin normal statistics are shown (Figures 22 and 23 respectively). MDS plots illustrate how closely related or different the samples are based on the features used in the SVM classification prediction. The statistics performed by Biodesix, Inc. can be found in Appendix A.

Statistical data analysis demonstrated that there are unique signatures of molecular changes that occur in the regions around the histological tumor border and that

there are aberrant features, resembling the tumor, in the normal microenvironment outside of the histological tumor border. The comparison between far tumor and near margin tumor indicated that there were no significant changes between them. Several possibilities may explain this result. The first possibility is that there are simply no differences between the mature tumor and the leading edge of the tumor, though this is unlikely. It may be that the changes are subtle enough that the differences are not detected. It is possible that the region considered "far tumor" in these samples was not far enough away from the growing edge of the tumor to detect differences. Lastly, the subtlety of these differences may require a larger sample pool to detect these changes.

In contrast to the far tumor and near margin tumor statistics, there were several features that differentiated the regions of far tumor and far normal. Features showing the best classification accuracy are highlighted in bold in Table 2. Feature-based classification prediction accuracy using these features was 91%, indicating the power of as few as six features to differentiate between tumor and normal. The comparison between the near margin tumor and the near margin normal revealed several differences, though consistent with results between far tumor and far normal. For example, features that were under-expressed in the near margin tumor were also under-expressed in the far tumor. No combination of features from this group was able to more accurately classify the two groups than the top individual feature classifier.

The purpose of comparing the far normal region to the near margin normal region was to determine if those differentiating features were also identified as significant in the far tumor and far normal comparison. Results of this analysis revealed that many features that are down in the near margin normal as compared to the far normal are also down in

the far tumor as compared to the far normal. Features determined over-expressed in tumor were also determined as over-expressed in the near margin normal. Nine features, represented in bold in Table 4, were able to classify the samples based on their far normal and near margin normal regions with a prediction accuracy of 75%.

When tissue is not collected by the researcher, it is difficult to control whether or not the tissue adjacent to the tumor is all cortex or all medulla. Tissue specimens may consist of four different adjacent normal tissue patterns: tumor and adjacent cortex, tumor and adjacent medulla, tumor and adjacent cortex followed by medulla, or tumor and adjacent medulla followed by cortex. To ensure that the tissue immediately adjacent to the kidney, whether it was cortex or medulla, had little influence on the statistics, a far normal and near margin normal analysis was performed. Results of the analysis showed that most features were the same within both groups, validating the grouping of all 34 samples independent of their near margin tissue type [Appendix B].

Table 1: Summary of patient information. W: White, B: Black

Age	**# Patients/Samples**
> 60	20
40-59	12
< 39	2
Gender	**# Patients/Samples**
Female	11
Male	23
Tumor Stage	**# Patients/Samples**
I	16
II	2
III	9
IV	7
Tumor Grade	**# Patients/Samples**
I	4
II	17
III	10
IV	3
Race	**# Patients/Samples**
W	30
B	2
Other	0
Unknown	2

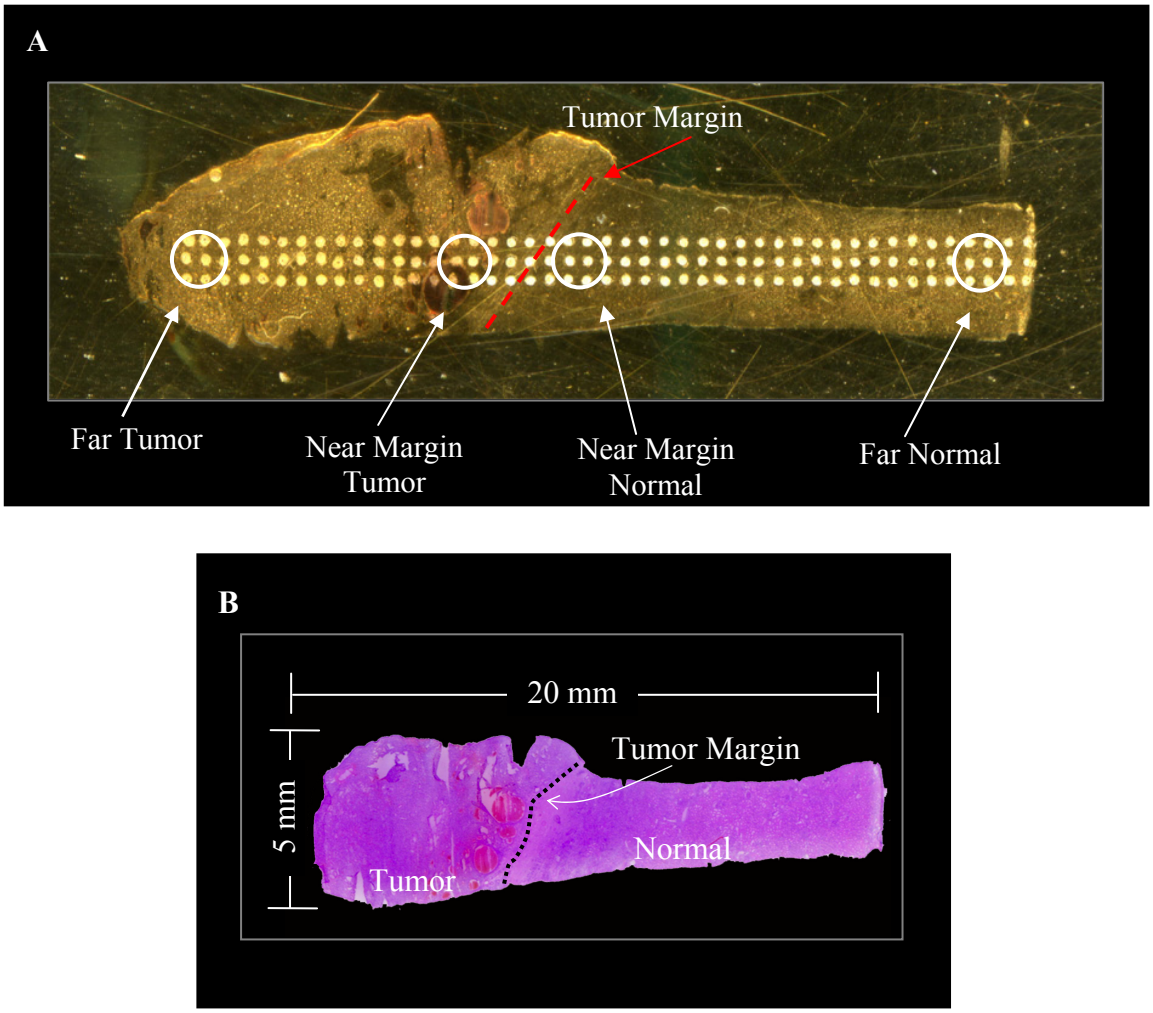


Figure 18: Regions of interest for statistical analysis of tumor margin profiles. A.) Optical image of section on MALDI plate with regions of interest marked; B.) Optical image of an H&E stained section marked by a pathologist.

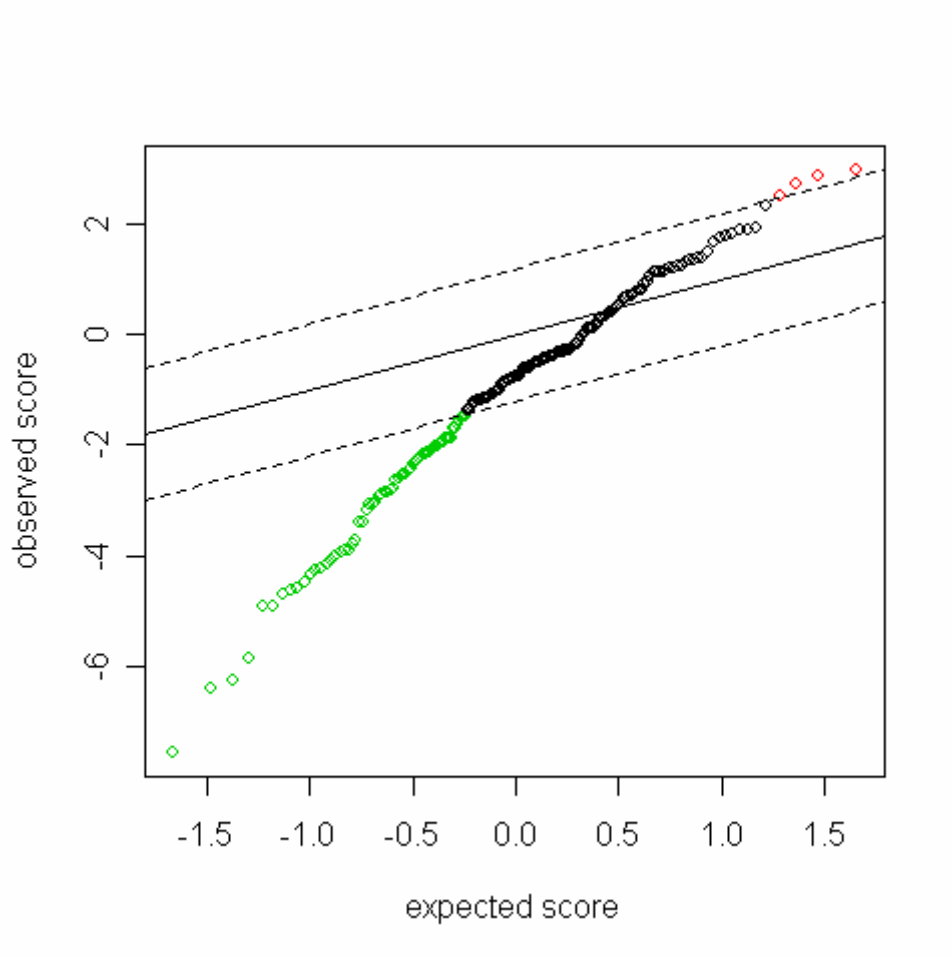


Figure 19: SAM statistic plot of results of far tumor versus far normal tissue. Red circles indicate significant features that are over-expressed in tumor. The green circles indicate features under-expressed in the tumor. Dotted lines represent the threshold (Δ) based on an FDR <0.01. Features have been arranged by their degree of difference in expression. Points to the right (top) or left (bottom) of the first point outside of the dotted line are called significant.

Table 2: Top differentially expressed features as determined by SAM and the permutation t-test in far tumor versus far normal regions. Data is presented with respect to the tumor. D: under-expressed in tumor; U: over-expressed in tumor. Feature class prediction values are also listed. Values listed had a bootstrap count of 99-100 out of 100 times. Values in bold indicate features used for classification. Feature: significant m/z value

m/z	Class Prediction	Expression
6717	0.88	D
9368	0.88	D
5351	0.87	D
5933	0.87	D
9616	0.84	D
4888	0.81	D
8649	0.81	D
6658	0.78	D
8350	0.78	D
12500	0.78	D
5260	0.76	D
9239	0.76	D
10296	0.76	D
10610	0.76	D
12275	0.76	D
5557	0.75	D
10089	0.75	D
10259	0.75	D
5363	0.74	D
5493	0.74	D
9514	0.74	D
6322	0.72	D
8577	0.72	D
8714	0.72	D
20922	0.72	D
6831	0.71	D
6946	0.71	D
8939	0.71	D
10275	0.71	D

m/z	Class Prediction	Expression
5039	0.69	D
6438	0.69	D
6572	0.69	D
8958	0.69	D
4275	0.68	D
5048	0.68	D
6429	0.68	D
8017	0.68	D
5151	0.66	D
8085	0.66	D
9071	0.66	D
13440	0.66	D
8562	0.65	D
13423	0.65	D
4670	0.62	D
4040	0.6	D
8769	0.6	D
10739	0.6	D
7862	0.59	D
9183	0.59	D
8544	0.56	D
14695	0.56	D
6113	0.54	D
6644	0.54	D
5648	0.44	D
5530	0.69	U
22198	0.63	U
4931	0.6	U
5165	0.49	U

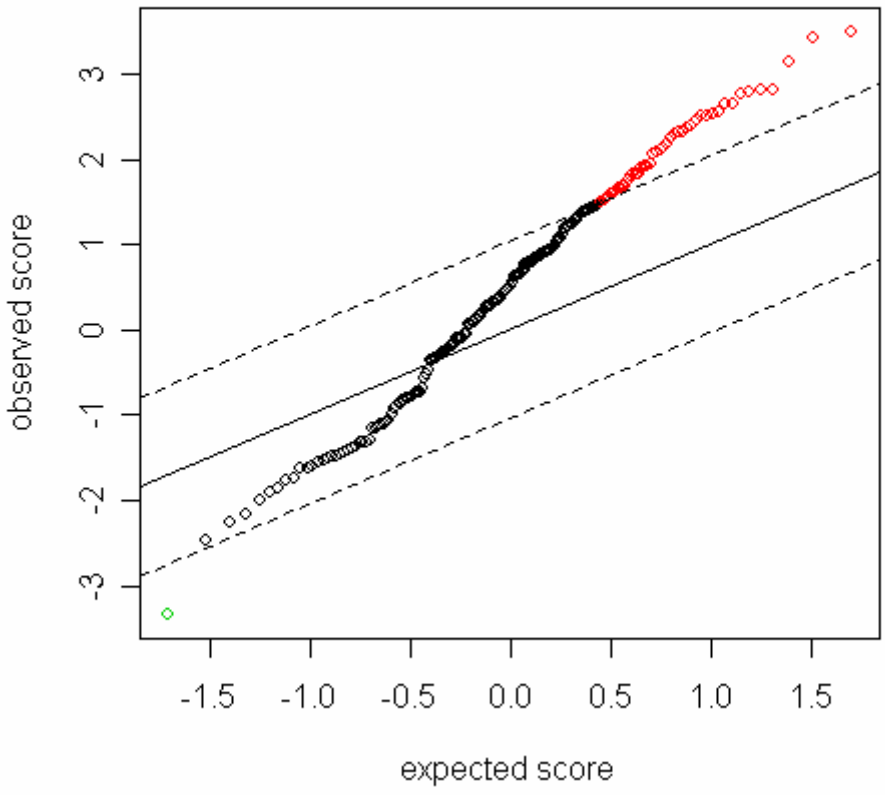


Figure 20: SAM statistic plot of results of near margin tumor versus near margin normal tissue. Green circles indicate features significantly over-expressed in near margin tumor. The red circles indicate features under-expressed in the near margin tumor.

Table 3: Top differentially expressed features as determined by SAM and the permutation t-test in near margin tumor versus near margin normal regions. Data is presented with respect to the near margin tumor. D: under-expressed; U: over-expressed. Feature class prediction values are also listed. Values listed had a bootstrap count of 99-100 out of 100 times. No combination of features was able to classify the groups better than the top individual classifier.

m/z	Class Prediction	Expression
10089	0.74	D
5260	0.72	D
5933	0.71	D
8649	0.69	D
9368	0.69	D
10296	0.69	D
5062	0.68	D
5039	0.66	D
6429	0.66	D
6946	0.66	D
5351	0.65	D
5557	0.65	D
5048	0.63	D
7862	0.63	D
20922	0.63	D
4275	0.62	D
4670	0.57	D
6831	0.56	D
6438	0.51	D
12345	0.65	U

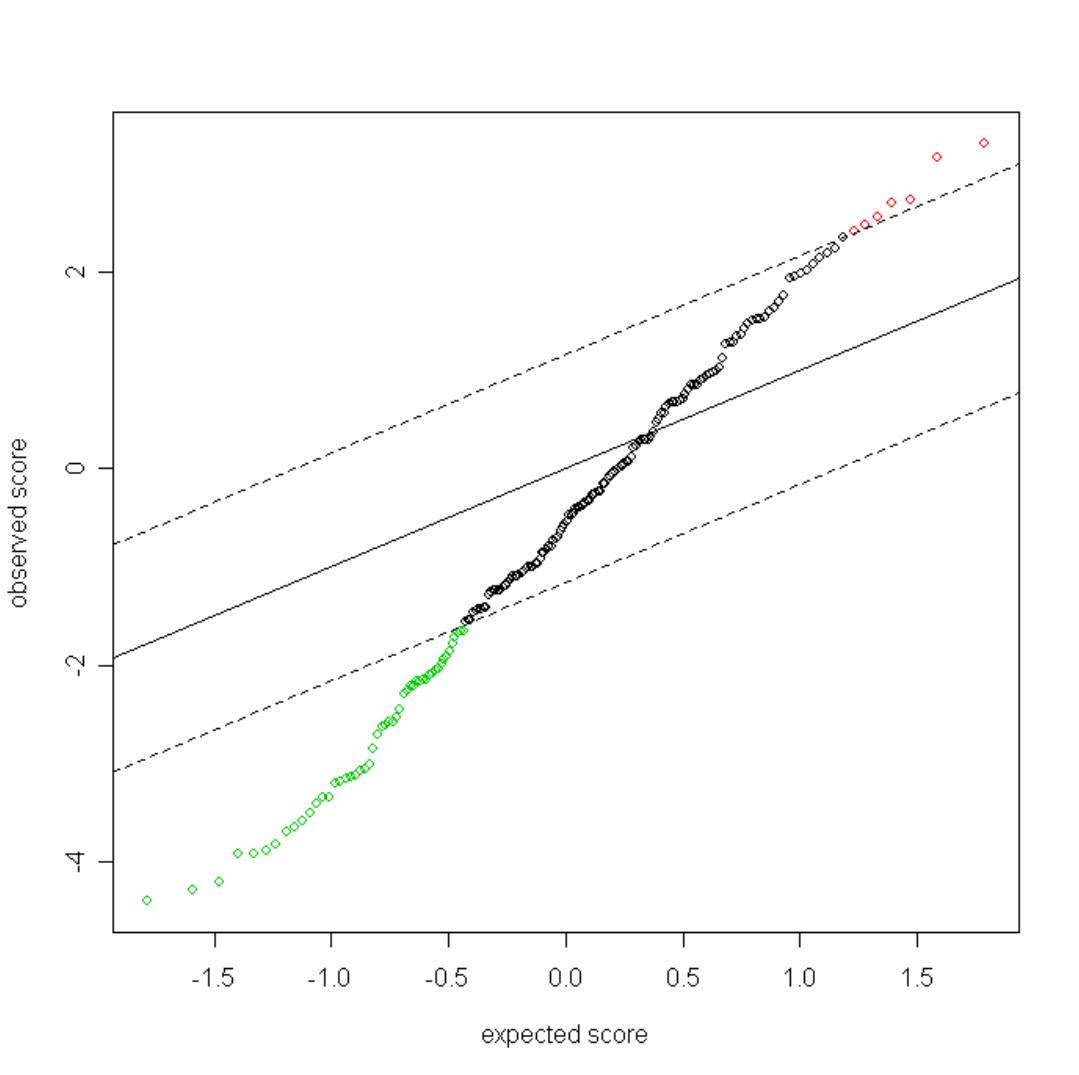


Figure 21: SAM statistic plot of results of near margin normal versus far normal tissue. Green circles indicate features under-expressed in near margin normal tissue. Red circles indicate features over-expressed in the near margin normal tissue.

Table 4: Top differentially expressed features as determined by SAM and the permutation t-test in near margin normal versus the far normal tissue. Data is presented with respect to the near margin normal. D: under-expressed; U: over-expressed. Feature classification prediction values are also listed. Values listed had a bootstrap count of 99-100 out of 100 times.

m/z	Class Prediction	Expression
6717	0.72	D
8649	0.72	D
9744	0.71	D
9953	0.71	D
4888	0.69	D
6429	0.69	D
5351	0.68	D
6658	0.68	D
10610	0.68	D
12275	0.68	D
12500	0.68	D
14695	0.68	D
5493	0.66	D
6438	0.66	D
8577	0.66	D
6322	0.65	D
8350	0.65	D
8958	0.65	D
9239	0.65	D
3319	0.63	D
5363	0.63	D
5933	0.63	D
9514	0.63	D
8017	0.62	D
10460	0.62	D

m/z	Class Prediction	Expression
6831	0.60	D
8238	0.60	D
9616	0.60	D
3895	0.59	D
4040	0.59	D
8562	0.59	D
9968	0.59	D
9368	0.57	D
9760	0.57	D
12767	0.57	D
20922	0.57	D
5310	0.56	D
8714	0.56	D
10834	0.56	D
19929	0.56	D
10259	0.50	D
8033	0.49	D
8085	0.49	D
4559	0.47	D
12345	0.47	D
3716	0.34	D
14217	0.66	U
5062	0.65	U
11347	0.65	U
5669	0.60	U

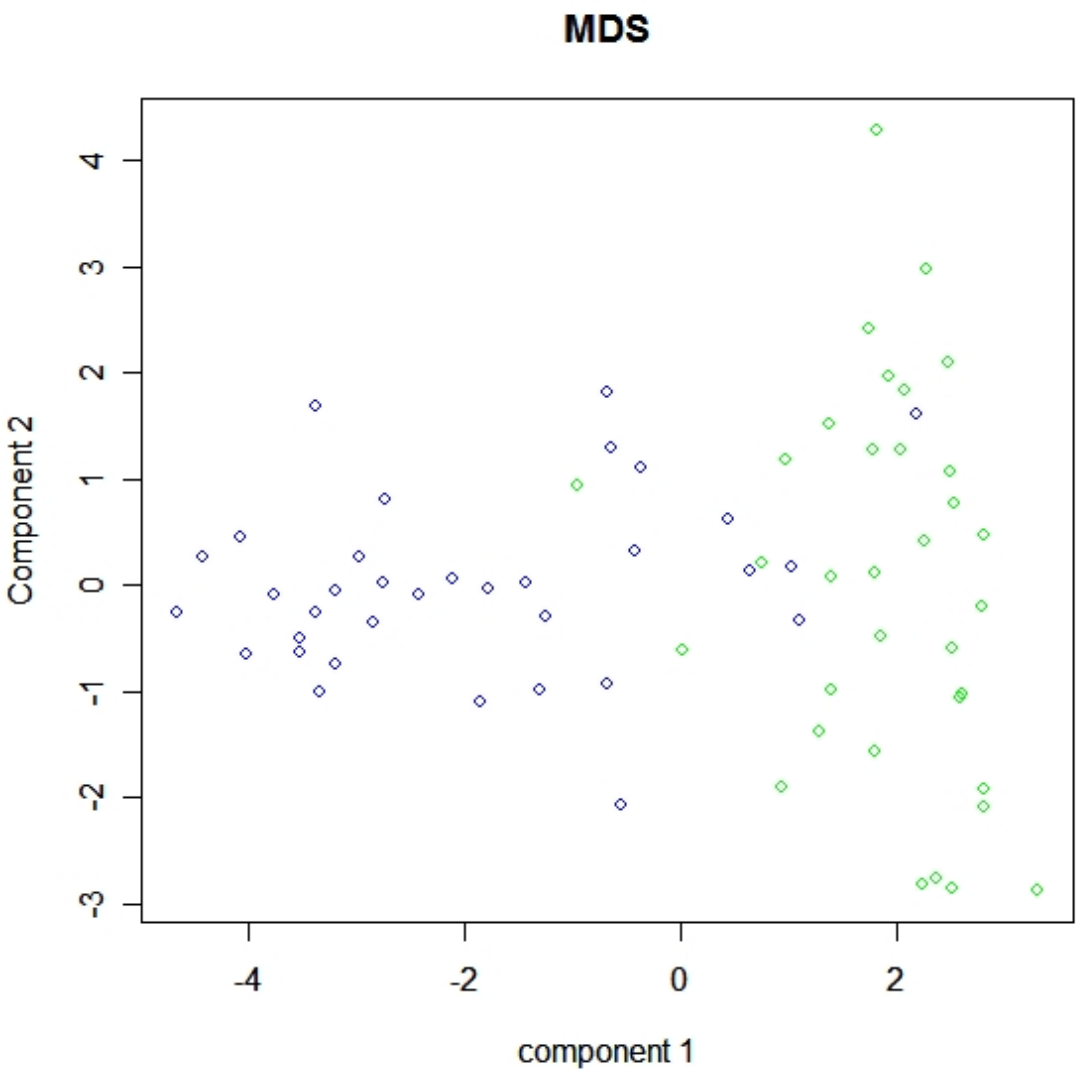


Figure 22: MDS plot of far tumor versus far normal. This plot demonstrates, in two dimensions, how closely related or how different the samples are based on features used for SVM class prediction. Blue circles represent normal samples and green circles represent tumor samples.

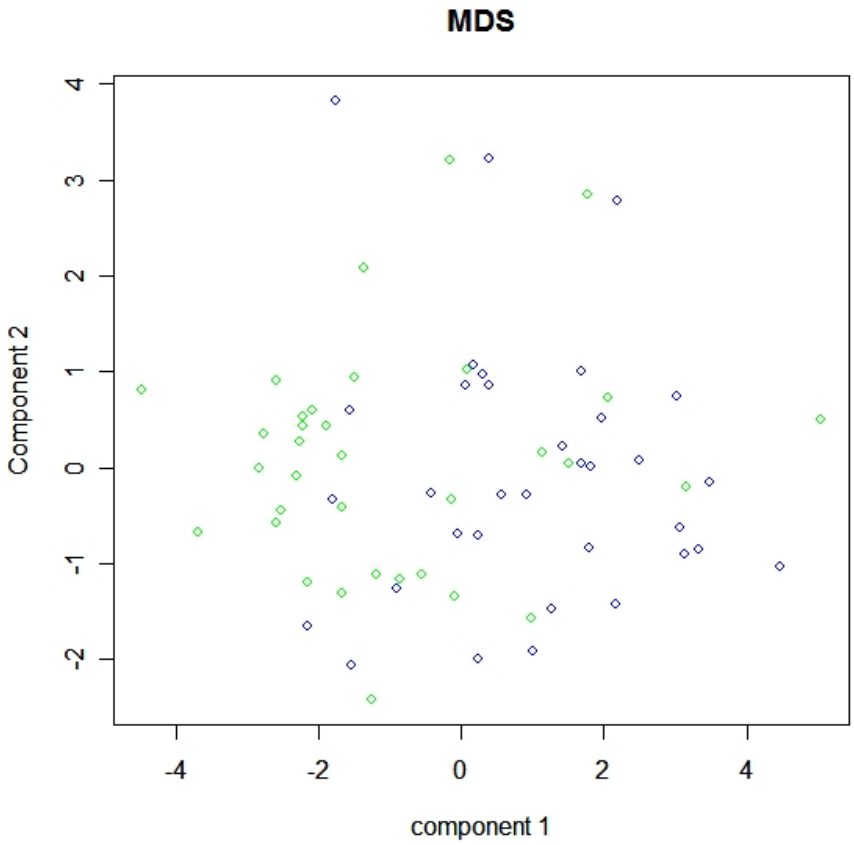


Figure 23: MDS plot of far normal versus near margin normal. This plot demonstrates how closely related, or different, the samples are in two dimensions based on features used for SVM class prediction. Blue circles represent far normal and green circles represent the near margin normal.

Pathology-Based Tissue Classification of MALDI MS Spectra

Once it was determined that there were features present outside of the histological margin that suggested a compromised environment, the second aim of this study was to determine if features that differentiate between tumor stage and/or grade could be associated with these tumor features that are present in the normal margin. Secondly, there is an immediate need for more reliable molecular markers of ccRCC to assist with current methods of pathological diagnosis of these tumors. Current diagnoses of ccRCC tumors relies on TNM staging, a definition of the anatomical location and spread of the tumor, and histology grading, which is based on cellular morphology. Both classification systems are associated with patient survival, yet the guidelines remain under review[2].

To perform statistical analysis, mass spectrometry data was processed, averaged by sample, and grouped into the following categories: non-tumor tissue (75 samples), tumor tissue (75 samples), high grade (grades III-IV), low grade (grades I-II), grade III, grade IV, high stage (stages III-IV), low stage (stages I-II), stage III, and stage IV. Samples were grouped independently of their acquisition source (Vanderbilt University (34 samples), Cooperative Human Tissue Network (3 samples), or Fox Chase Cancer Center Tumor Repository (38 samples). The sample information is summarized in Table 5. Using the statistics SAM and the permutation t-test, the following comparisons were made: non-tumor vs. tumor (pairwise) [Figure 24, Table 6], high stage vs. low stage [Figure 26, Table 7], stage III vs. stage IV [Figure 28], high grade vs. low grade [Figure 29], grade III vs. grade IV. Results reported in the tables represent features that were within an FDR of less than 0.01 for both statistic algorithms, unless otherwise noted.

Statistical analysis resulted in reliable separation of most of the comparisons performed. Analysis of the tumor and non-tumor groups resulted in a very distinctive separation. Out of the features determined as significant, as few as six features (m/z values 5351, 6717, 9616, 9368, 4931, 5178 and 12345) were able to classify these two groups with 95% accuracy. The MDS plot, based on these features, is shown in Figure 25.

Since the TNM staging system describes, with minimal subjectivity, the extent of tumor growth in and around the kidney, including metastasis, it was expected that high stage versus low stage would result in a more reliable separation than between stages III and IV. The comparison of high stage versus low stage resulted in several differential features. The class prediction accuracy of as few as four of these features (m/z values 7930, 7662, 4385, 4107) was 80%. The MDS plot of feature separation based on the values used for classification is shown in Figure 27. Only the SAM statistic [Figure 29] was able to determine features that differentiate between stages III and IV. Six m/z values were determined significant, but three of them were doubly-charged peaks of the remaining three, which corresponded to hemoglobin. Using all six m/z values, the classification accuracy was 62%, but when excluding the doubly-charged peaks, hemoglobin alone is incapable of distinguishing between stages III and IV ccRCC tumors. Given that there were at least 24 samples in each of the two groups, it is surprising that there were not more differential features. The inability to differentiate these two groups could indicate that there were not enough samples. The second possibility could be that the stage IV tumors are best separated into two groups, those that have distant metastases and those that do not. If samples were grouped into Stage III,

Stage IV no metastases, and stage IV with metastases, there may be a better separation of these groups; however, sample information obtained from the tumor bank facility where half of the samples were obtained lacked information on which stage IV samples had distant metastases and which did not.

Because of the high subjectivity involved in assigning tumor grades, it was not expected that there would be many differences between tumors when grouped according to their grade diagnosis. The comparison of high grade versus low grade resulted in three distinguishing features by the SAM statistic (m/z values 4040, 4107, and 8448) [Figure 29]. There were no features identified by the permutation t-test. The best classifier of these two groups was m/z 8448, with an accuracy of 71%. The MDS plot of these two groups using the feature m/z 8448 is shown in Figure 30. Neither the SAM nor the permutation t-test were able to identify differential features between grade III and grade IV. The inability to determine differential markers between grades III and IV recapitulates the clinical pathology of these tumors. Based on current histological methods, the grading of ccRCC tumors remains subjective, which may be the underlying influence in not being able to distinguish these two grades.

Table 5: Summary of patient information. W: white, B: black. The asterisk (*) represents overlap in the grading system, for example, one tumor might be graded as high grade and another as grade II-III.

Age	**# Patients/Samples**
> 60	48
40-59	23
< 39	4
Gender	**# Patients/Samples**
Female	24
Male	51
Tumor Stage	**# Patients/Samples**
I	17
II	9
III	25
IV	24
Tumor Grade	**# Patients/Samples**
I	6-7*
II	29-32*
III	26-31*
IV	8-10*
Race	**# Patients/Samples**
W	66
B	2
Other	1
Unknown	6

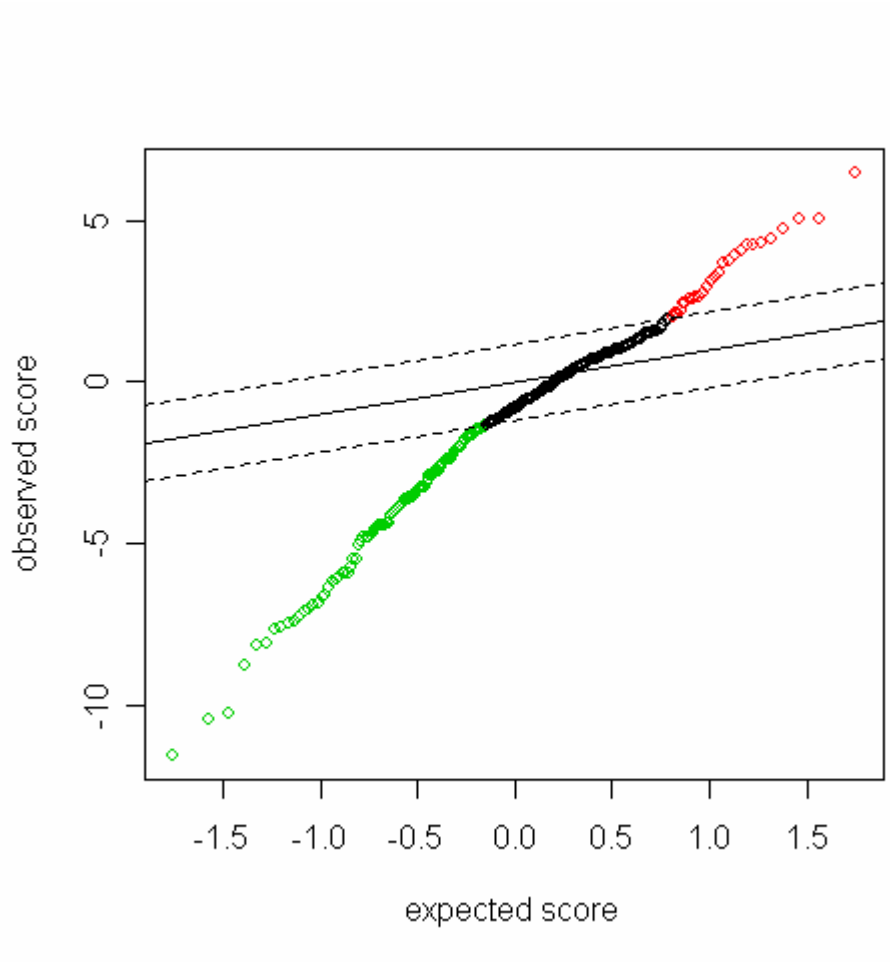


Figure 24: SAM plot of the pairwise comparison of tumor versus normal. Green circles represent features under-expressed in the tumor. Red circles indicate features over-expressed in the tumor.

Table 6: Top differentially expressed features as determined by SAM and the permutation t-test in non-tumor and tumor tissue. Data is presented with respect to the tumor. D: under-expressed; U: over-expressed. Feature classification prediction values are also listed. Values listed had a bootstrap count of 99-100 out of 100 times. Features in bold were used for the combined class prediction.

m/z	Class Prediction	Expression
5351	0.91	D
9616	0.88	D
9368	0.87	D
6717	0.86	D
5932	0.85	D
12272	0.85	D
6658	0.83	D
8652	0.83	D
10610	0.83	D
12500	0.83	D
4888	0.81	D
5362	0.81	D
8958	0.81	D
9239	0.8	D
6830	0.79	D
7799	0.78	D
8017	0.77	D
10258	0.77	D
6358	0.76	D
5494	0.75	D
5557	0.75	D
9514	0.75	D
10089	0.75	D
14692	0.75	D
5260	0.74	D
6322	0.74	D
7174	0.74	D
8577	0.74	D
5048	0.73	D
6946	0.73	D
9070	0.73	D

m/z	Class Prediction	Expression
9456	0.73	D
20921	0.73	D
20937	0.73	D
4275	0.71	D
5203	0.71	D
6073	0.71	D
6438	0.71	D
8350	0.71	D
8939	0.71	D
10275	0.71	D
5310	0.7	D
6106	0.7	D
8714	0.7	D
3341	0.69	D
4284	0.69	D
5038	0.69	D
5647	0.69	D
6114	0.69	D
6430	0.69	D
6568	0.69	D
9761	0.68	D
4041	0.67	D
5151	0.67	D
9970	0.67	D
6312	0.67	D
10834	0.67	D
8104	0.67	D
4670	0.66	D
6775	0.66	D
8562	0.66	D
8769	0.66	D

Table 6, continued.

m/z	Class Prediction	Expression
3808	0.65	D
6179	0.65	D
7205	0.65	D
9953	0.65	D
4679	0.64	D
8237	0.64	D
7343	0.63	D
9334	0.63	D
10296	0.63	D
13439	0.63	D
9744	0.63	D
4343	0.62	D
10739	0.62	D
9183	0.62	D
5852	0.62	D
8544	0.62	D
13423	0.6	D
3030	0.59	D
5656	0.59	D
4218	0.59	D
7372	0.59	D
8181	0.59	D
13029	0.59	D
3814	0.59	D
8087	0.58	D
4663	0.58	D
3193	0.57	D
2815	0.57	D
2929	0.57	D
3187	0.56	D

m/z	Class Prediction	Expression
4431	0.55	D
16792	0.54	D
6644	0.53	D
4128	0.51	D
21751	0.51	D
8448	0.46	D
3848	0.35	D
7266	0.27	D
4931	**0.74**	U
5178	**0.74**	U
12345	**0.71**	U
5530	0.69	U
4385	0.68	U
11071	0.67	U
14092	0.67	U
21265	0.67	U
4107	0.66	U
7779	0.66	U
11651	0.66	U
4114	0.65	U
6875	0.65	U
11388	0.65	U
22200	0.65	U
11969	0.64	U
20198	0.63	U
4304	0.62	U
4730	0.61	U
11986	0.6	U
2951	0.59	U
22777	0.59	U
15333	0.57	U
4514	0.56	U

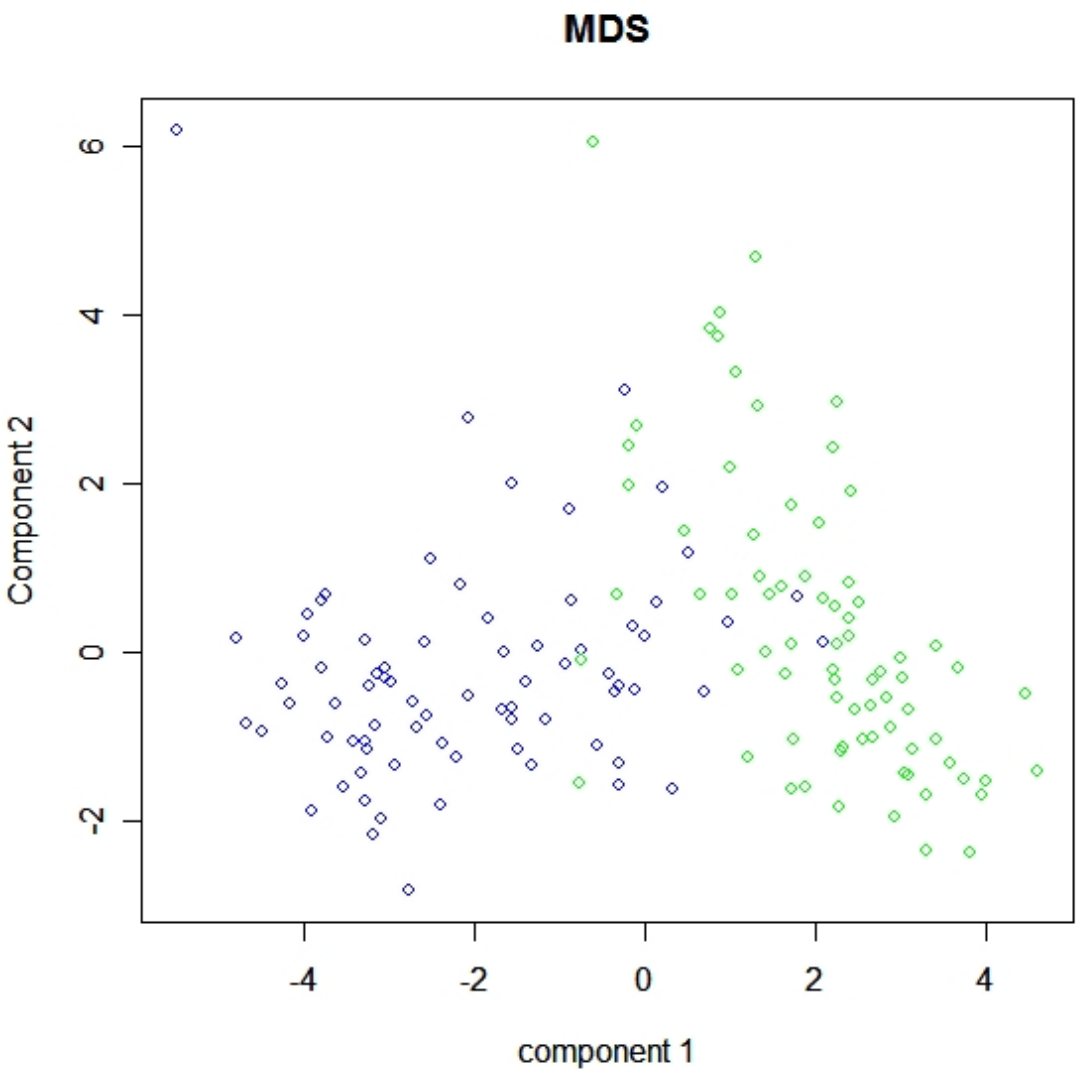


Figure 25: MDS plot of tumor versus non-tumor tissue. The plot represents the tissue sample separation based on the features used in SVM class prediction. Green circles indicate tumor tissue samples and blue circles indicate non-tumor tissue samples.

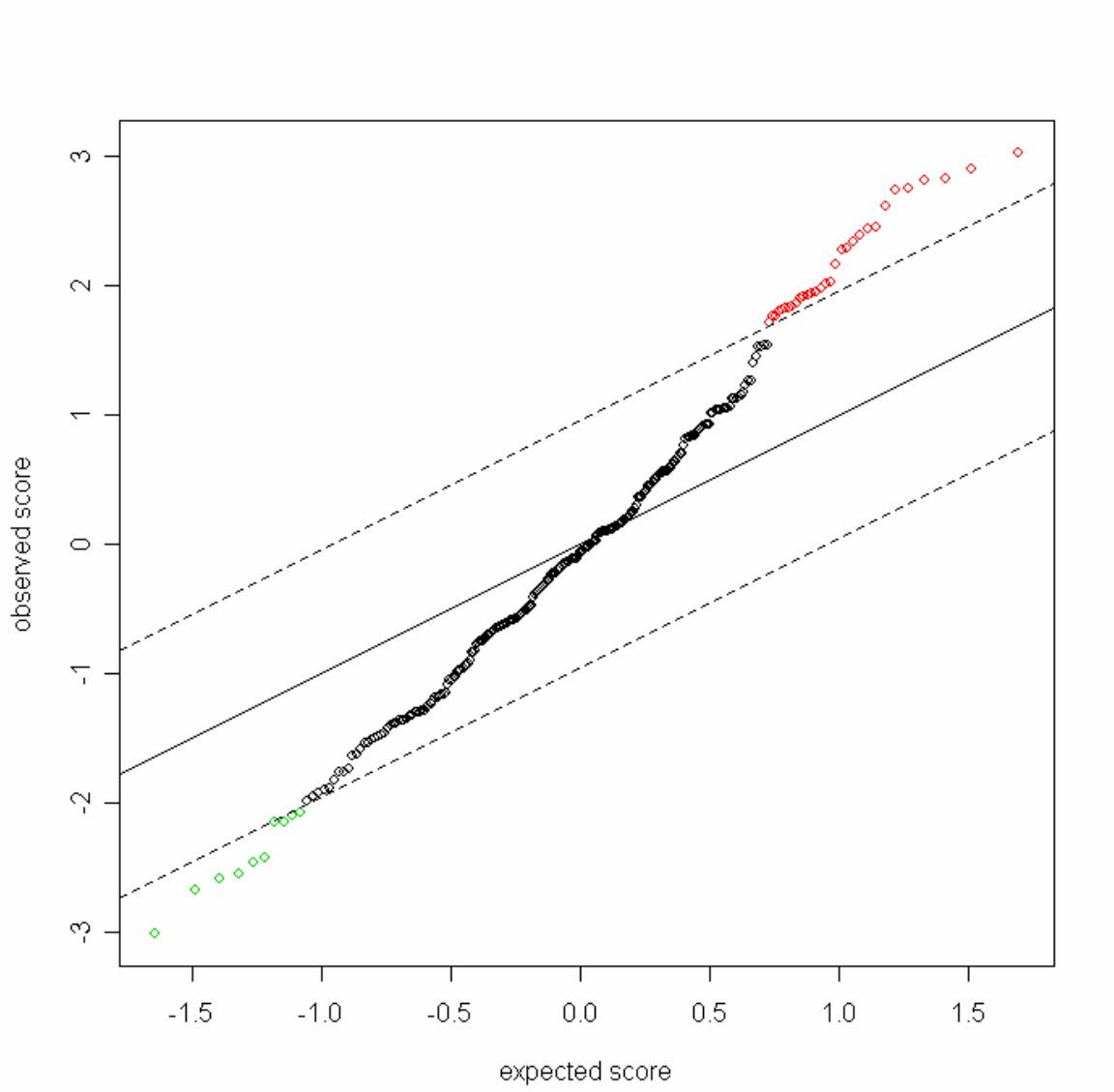


Figure 26: SAM plot of high stage versus low stage. Green circles indicate features under-expressed in the high stage tumors, whereas red circles indicate features over-expressed in the high stage tumors.

Table 7: Top differentially expressed features as determined by SAM and the permutation t-test in low stage (I-II) versus high stage (III-IV) tumors. Data is presented with respect to the high stage tumors. D: under-expressed; U: over-expressed. Feature classification prediction values are also listed. Values listed had a bootstrap count of 99-100 out of 100 times. Features in bold were used for class prediction.

m/z	Class Prediction	Expression
7560	0.73	D
7662	0.72	D
7930	0.76	D
8136	0.6	D
15126	0.56	D
4730	0.76	U
4385	0.75	U
3420	0.72	U
4107	0.72	U
8448	0.65	U
4613	0.64	U

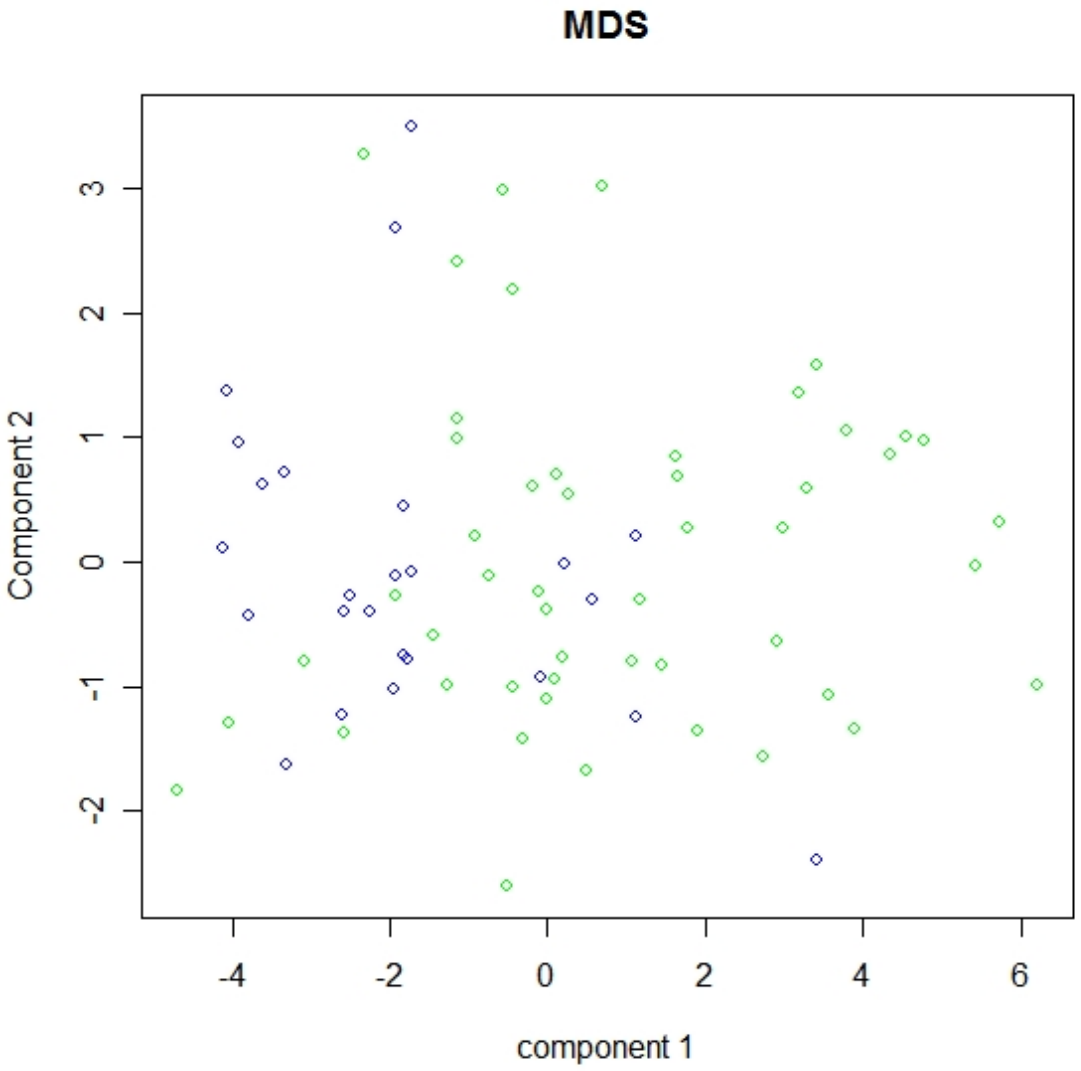


Figure 27: MDS plot of high stage versus low stage. The plot represents the tissue separation, in two dimensions, based on the features used in SVM class prediction. Green circles indicate high stage tumors and blue circles indicate low stage tumors.

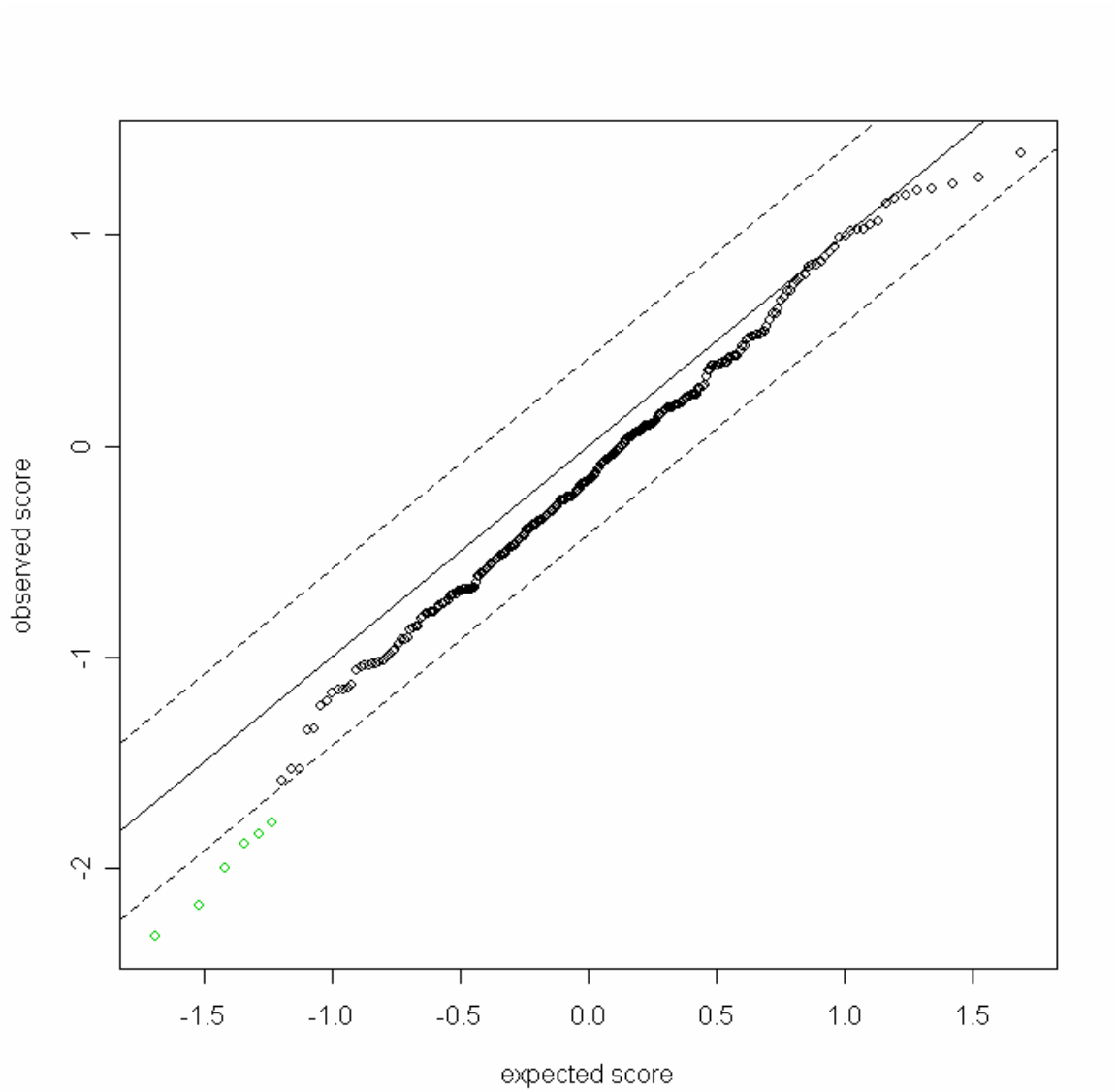


Figure 28: SAM plot of stage III versus stage IV. Green circles indicate features that are significantly under-expressed in stage IV tumors as compared to stage III tumors.

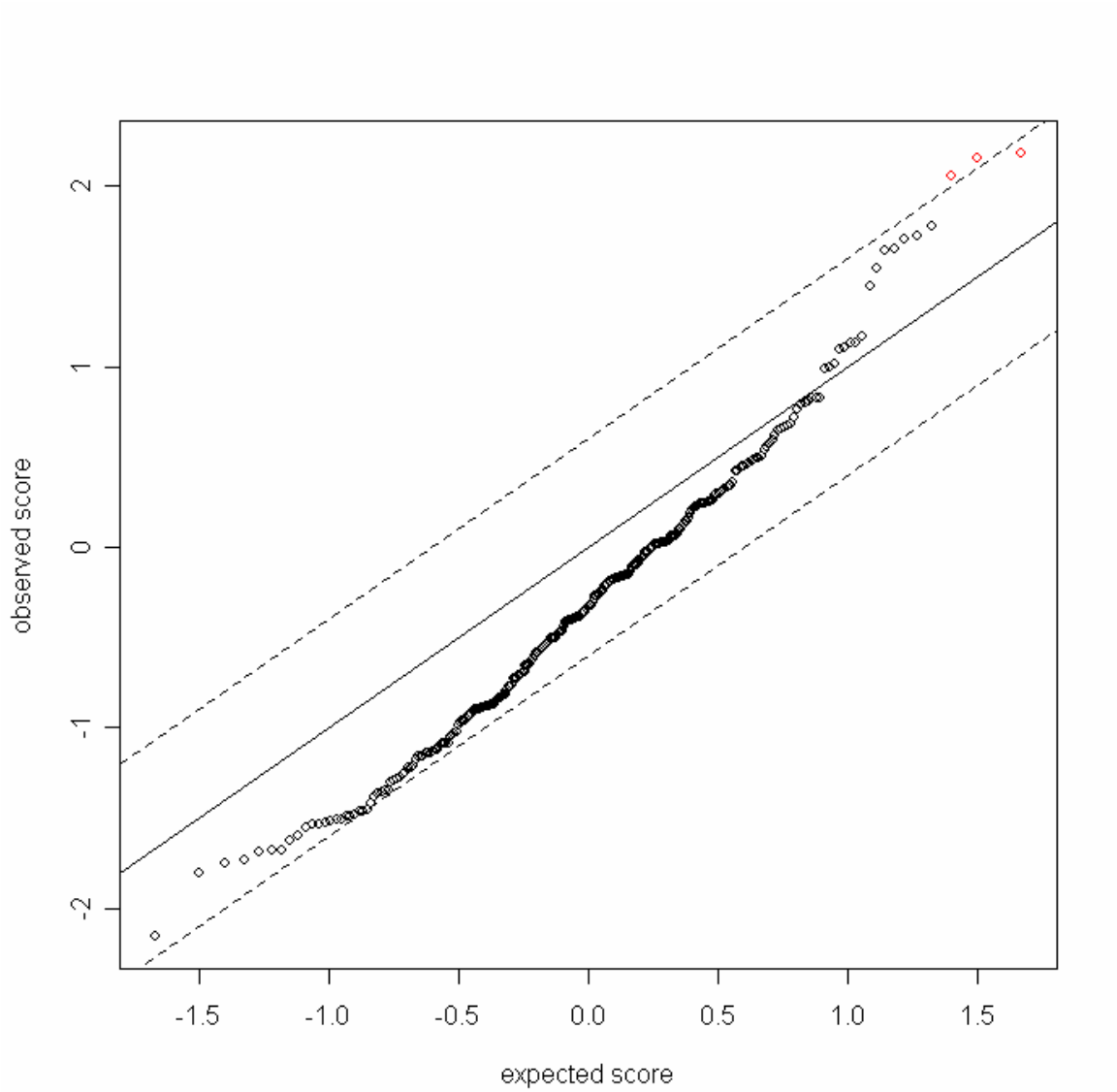


Figure 29: SAM plot of high grade versus low grade. Red circles indicate features that are over-expressed in high grade tumors.

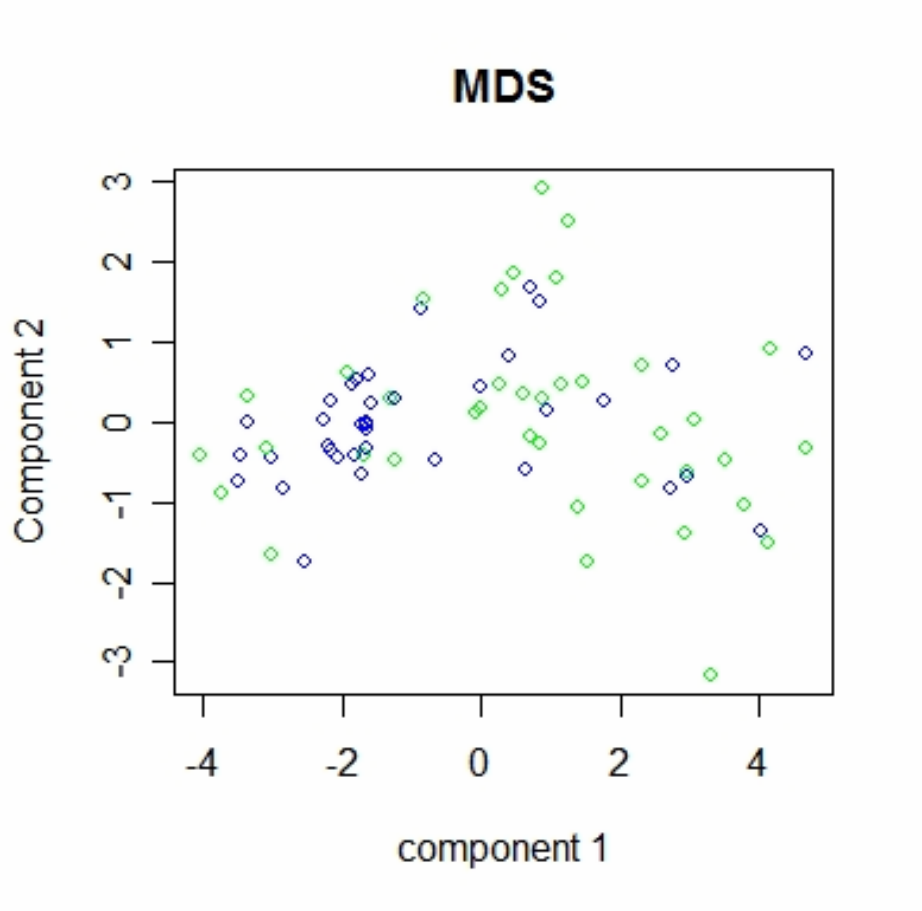


Figure 30: MDS plot of low grade and high grade tumors. Green circles represent high grade tumors and blue circles indicate low grade tumors. The MDS plot is based on features used in class prediction.

Assessing Patterns of Molecular Change in Tumor Margins

The final objective of this project was to assess how features change in and around the tumor margins and how this pattern relates to tumor aggressiveness. It was expected that some features would show a correlation to the histological border, while others would indicate changes occurring beyond the histological border, namely the features whose expression patterns resembled their counterparts in the tumor. A schematic of potential trends is seen in Figure 31. As with most tumors, there are no molecular markers for tumor aggressiveness in renal cell carcinoma, thus the attempt was made to determine if these patterns correlate with the current standards of clinical pathology diagnoses, which are Fuhrman grading and TNM staging.

The mass spectra were collected in a way that allows for the mapping of feature expression from tumor into adjacent normal tissue. As implied, only the samples containing tumor and attached adjacent normal were considered. Because visible borders that define tumor margins do not follow a straight line, the analysis of feature patterns requires that the distance of each matrix spot to the margin be measured accurately. In doing so, a feature expression can be mapped according to its distance from the histologically-defined tumor margin. To obtain the distance, the optical image of a hematoxylin and eosin stained section was used to define the tumor border on the optical image of the spotted tissue. A custom program was designed to calculate the distance of the matrix spot to the closest point along the defined histological margin. A spreadsheet was made containing the distance of each matrix spot and its corresponding mass spectrometry file.

To best represent the data, the feature amplitude (log2 scale) from each matrix spot was plotted against its distance to the histological margin. A locally-weighted-regression scatter-plot smoothing (LOWESS) curve was laid over the scatter plot to show the trend of change of feature amplitude across the histological margin. The first derivative of the LOWESS trend line was plotted to show the position where the maximum rate of change of feature amplitude occurs. An example of this method is shown in Figure 32.

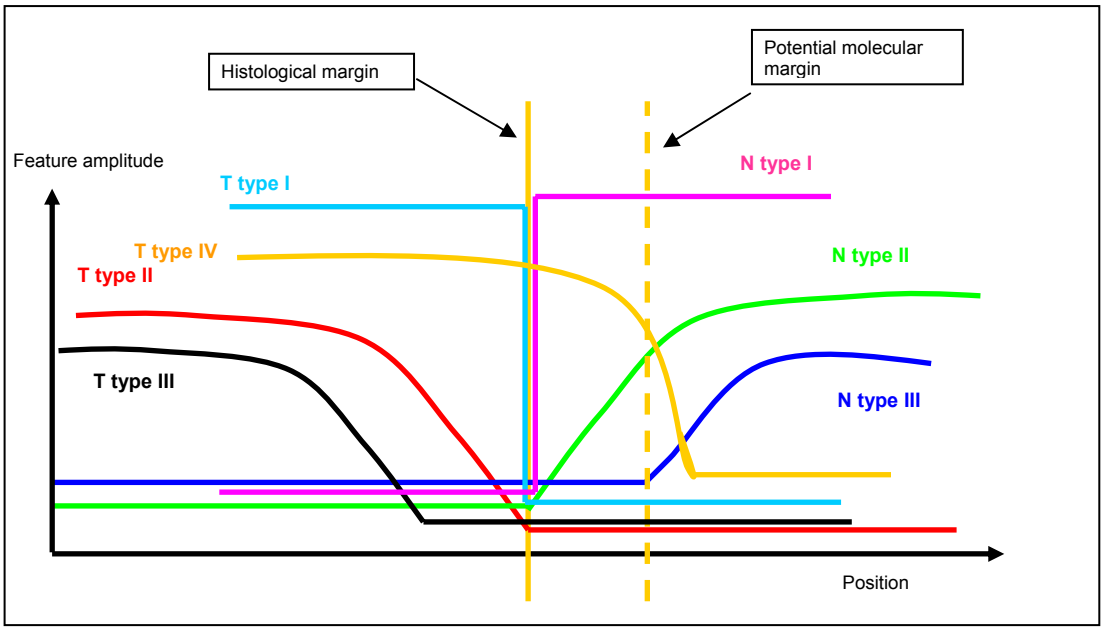


Figure 31: Potential patterns of molecular features traversing the histological tumor margin. T type I defines features over-expressed in tumor that follow the histological margin with an immediate change in feature expression; N type I defines features under-expressed in tumor that follow the histological margin and immediately change in feature expression; T and N type II describe features that begin changing at the margin. T and N type III and T type IV define features that begin changing before or after the histological margin.

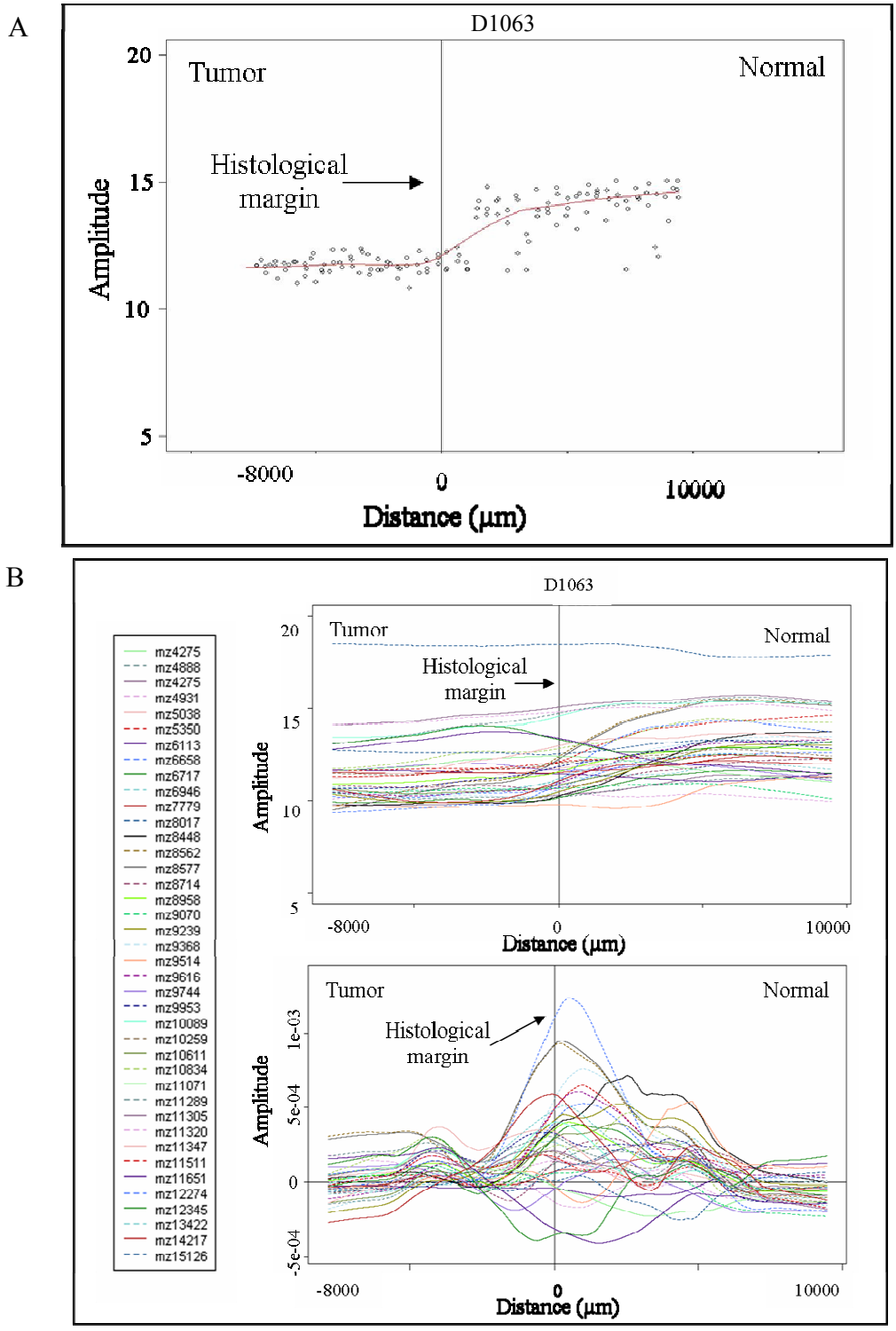


Figure 32: Demonstration of margin plot analysis. A.) LOWESS fit line of scatter plot data. B.) Top: LOWESS line of 40 selected significant features; Bottom: First derivative of all LOWESS lines (x-axis error approximately +/- 400 µm).

As described in Figure 31, results were expected to show features that define the histological margin and features that change before or after the histological margin. For broad interpretation, 40 significant features were plotted for each of the 24 patients (Appendix C). These features represented features that were determined as under-expressed and over-expressed in the ccRCC tumors as well as aberrant molecular features present outside of the histological margin. A careful assessment of the data included reviewing both the LOWESS and first derivative lines. First derivative lines were auto scaled in the y-axis direction for ease of reviewing the data. However, in some cases, this enhanced the noise. Secondly, some fluctuations in the first derivative result from slight biological variations in the LOWESS line. Upon assessment of the plots, no "T type I" or "N type I" trends were observed, instead, there were strong indications that "T type II", "N type II" and "N type III" trends were present. These observed trends are summarized in Figure 33.

The final objective of this study was to determine if molecular expression patterns traversing the histological tumor margin correlated with pathology diagnoses. Molecular features of interest included those that were differential between tumor and normal as well as those indicating aberrant characteristics in the normal-appearing tissue outside of the tumor margin. After observing the LOWESS trends [Figure 32 and Appendix C] and reviewing the statistics, the features chosen to assess these patterns were m/z values 4888, 5351, 6717, 8577, 9368, 10611, and 12274 [Appendix D].

Interestingly, depending on the patient, these features exhibited strong patterns of "N" types II and III. Multiple characteristics of the plot and of the patient information were considered in order to find the best trend in the data. The plot characteristics

monitored were the point at which the maximum rate of change occurred and the point at which the feature amplitude began to plateau. Patient information included tumor grade, tumor stage, and the TNM tumor size designation. A summary of observations is listed in Table 8.

There was not a clear correlation between the pathological diagnosis and the molecular pattern of expression through the tumor margin. For the features focused on, most samples followed an N type III pattern. There was a group of samples in which margin trends could not be assessed. This group includes four high stage tumors and three low stage tumors as well as three high grade tumors and four low grade tumors. Overall, this group may be comprised of two sets of samples. The low stage/grade tumors may not present changes because of their low malignancy. Sample ID "D1345" may be an example of the observations in the high stage tumors. As seen in the chart, sample "D1345", a high grade and high stage tumor, does not plateau until at least 15000 micrometers past the histological margin. Sample "D529a" was sampled up to only 5000 micrometers past the histological margin. Although "D1351" had tissue sampled up to 10000 micrometers, it was difficult to decipher and the LOWESS lines appeared to be relatively flat. Sample "D1925" was also difficult to decipher, with the fluctuations on the tumor side of the margin appearing to be biological. The flatness of the LOWESS and minimal change in the first derivative seem to indicate that the actual molecular change occurs beyond the 10000 micrometer distance.

These observations suggest that a distinct pattern exists that may be indicative of a molecular margin. With respect to the tumor, features began to change at the margin or after the margin. Cases in which features begin to change after the margin would be

indicative of more severe tumor invasion. This was observed in several samples. There were also samples with no observational change. These may be indicative of two possible scenarios. The first scenario includes samples with very low malignant tumors that do not yet express these distinct patterns of change. For those samples, there may be more specific features that best describe their patterns of change. These features could be indicative of less invasive tumors. The second scenario includes samples that are highly malignant and their molecular margin lies beyond the tissue available for the analysis. The most severe molecular margin existed 15 mm beyond the visually defined tumor border and this was the longest extent of adjacent normal tissue analyzed. These results highlight the complexity of molecular events that occur around the histological tumor margin. Most importantly, the results show that tissue as far as 15000 micrometers away from the tumor margin can be compromised by the presence of tumor and go unnoticed under current histological analyses.

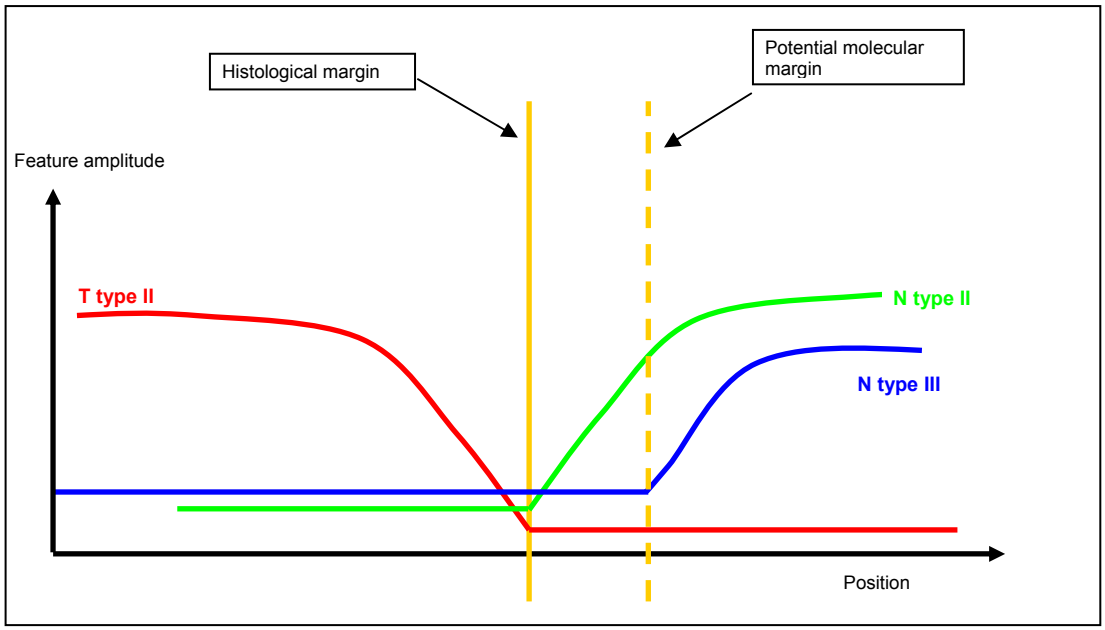


Figure 33: Observed molecular patterns traversing the tumor margin. The dotted lines represent features that begin changing before the margin and gradually increase or decrease past the margin.

Table 8: Table summarizing observed margin plot trends. Size: the size of the tumor based on TNM staging; Start: the point at which feature amplitude begins to change; Max Change: refers to the maximum point of change; Plateau: the point at which the feature amplitude is no longer changing; N/A: distance could not be assessed.

Sample ID	Grade	Stage	Size	Max Change (μm)	Plateau (μm)
D1349	2	1	1	0	2500
D0760	4	3	2	0	1100
D1930	2	1	1	800	2200
D0767	1	1	1	1250	7500
D1923	2	3	3	1250	4800
D0530	2	1	1	1250	3000
D1063	2	2	2	1250	4000
D1067	3	3	3	1500	2200
D1922	2	1	1	1800	3800
D0902	2	3	3	1800	4000
D1341	1	1	1	2000	4000
MAD060049	2	2	2	2400	6000
MAD060044	2	1	1	2500	4500
D1062	2	1	1	3000	7500
D1344	2	4	1	4500	6000
D1340	2	4	3	5000	6000
D0901	3	4	4	5000	9000
D1345	4	3	3	11000	15000
D1347	2	4	1	N/A	N/A
D0764	1	1	1	N/A	N/A
D1064	3	1	1	N/A	N/A
D1920	2	1	1	N/A	N/A
D0529a	2	3	3	N/A	N/A
D1351	3	3	3	N/A	N/A
D1925	3	4	3	N/A	N/A

Identifying Differentially Expressed Biological Features

To identify significant proteins, three methods were used, as summarized in Figure 34. In method A, the proteins from homogenized tissue were digested and loaded onto immobilized pH gradient gel strips for isoelectric focusing of the peptides. After focusing, the IPG strips were cut into 15 pieces. Peptides were extracted from the gels and desalted with C_{18} resin using a stepwise elution gradient of 20%, 40%, 60% and 80% acetonitrile. After desalting, peptides were analyzed by LC-MS/MS on an ion trap mass spectrometer. Liquid chromatography was performed on a C_{18} capillary column coupled to the mass spectrometer. Data dependent scanning was used to obtain MS and MS/MS spectra of peptides for sequencing. A database search with Sequest was performed to match potential proteins to their theoretical sequence. In method B, intact proteins were separated by reversed-phase liquid chromatography on a C_8 column and fractions were collected in a 96-well PCR plate at one minute intervals. To identify wells containing proteins of interest, each well was analyzed by MALDI-TOF MS. Fractions containing the m/z values of interest were digested with trypsin and analyzed with LC-MS/MS as described above. Method C utilized the additional separation power of 1D PAGE to further separate proteins in the fractions of interest from LC, as above. After gel separation and visualization, bands of interest were cut out. Following in-gel digestion with trypsin, resulting peptides were then sequenced by LC-MS/MS as previously described. After the database searches, ProteinProphet software was used to determine the probability that a protein had been correctly identified based on the available peptide sequence[152].

When identifying proteins, emphasis was placed on the features determined as indicative of aberrant environments outside of the histological margin. Table 9 lists the proteins identified by these methods and those previously identified. From this study 14 proteins were identified. Features identified that were over-expressed in tumor include calpactin I, calgizzarin, MIF (macrophage migration inhibitory factor), and thymosin beta-10. Features under-expressed in tumor were identified as calcyclin, calbindin, cytochrome c, NADH-ubiquinone oxidoreductase MLRQ subunit, a ubiquinol-cytochrome c reductase complex protein, and several cytochrome c oxidase polypeptides (VIC, VB, VIIA2, VIIC, VIII2). The proteins that indicated a compromised environment outside of the histological margin were cytochrome c, the cytochrome c oxidase polypeptides, and the NADH-ubiquinone oxidoreductase MLRQ subunit.

Several of the proteins identified are members of the S100 family of proteins. These proteins include calpactin I (S100-A10), calgizzarin (S100-A11), calbindin (S100-G), and calcyclin (S100-A6). The S100 protein family is comprised of calcium-binding proteins with two helix-loop-helix EF-hand motifs. There are approximately twenty human S100 family members, sixteen of which (S100A1-S00A16) are clustered on the chromosome 1q21 region, where a number of chromosomal abnormalities occur in neoplasms. This protein family is responsible for a variety of cellular processes including cell proliferation and differentiation and intracellular calcium regulation. Results show an increased expression of Calgizzarin in clear cell renal carcinoma. Calgizzarin, the S100A11 protein, plays a potential role as a tumor suppressor and its overexpression in renal tumors has been observed[153]. In this study, Calpactin I was also over-expressed in tumor cells. This is consistent with previous studies of renal cell carcinoma. The S100-

A10 protein forms cellular ligands with annexin II and this complex is known to activate plasminogen to form plasmin. The loss of S100A10 from the extracellular surface of cancer cells results in a significant loss of plasmin generation. Therefore it is possible that the overexpression of S100A10 and annexin II in RCC would be correlated with poor prognostic indicators[154].

Protein S100-G, or calbindin, is a vitamin D dependent calcium binding protein. Vitamin D is a factor that inhibits cell proliferation and promotes apoptosis. These functions are mediated by the vitamin D nuclear receptor, which is present in healthy kidneys. Its binding to DNA is required for proper function. In clear cell renal carcinoma, its DNA binding is altered. Although vitamin D expression levels do not change in renal tumors, it is expected that decreased vitamin D binding activity will cause abnormal activities of vitamin d-downstream genes, such as calbindin[155].

Calcyclin is a S100A6 calcium binding protein whose activity has been implicated in numerous molecular functions including tumorigenesis. Increased levels of calcyclin have been associated with decreased metastasis and inhibition of cell migration and to anchorage independent growth in human osteosarcoma[156]. Decreased expression of calcyclin has been observed in prostate cancer and its precursors. The mechanism of its decreased expression is unknown, yet some suggest that it may involve methylation of CpG sites within the promoter region[157].

Thymosin beta-10 has also been shown to be over-expressed in human renal cell carcinoma[158]. This protein is a member of the thymosin family, which is composed of three classes, alpha, beta and gamma. The beta class plays a role in cellular structure. Thymosin beta-10 expression is related to cell growth in proliferating tissues and its

expression has been reported in tumors. In tumorigenesis, beta thymosins bind and sequester G-actin, which plays a role in maintaining cellular structure. It is believed that elevated thymosin beta-10 levels in malignant tumors may be correlated with actin polymerization. It is known that overexpression of exogenous thymosin beta-10 causes a disassembly of actin stress fibers. Disassembly of the stress fibers enhances cell migration, which enhances conditions for malignant growth[159].

Macrophage migration inhibitory factor (MIF) has been implicated in regulating tumor migration and expression of angiogenic factors in hepatocellular carcinoma as well as in the regulation of host inflammatory and immune responses[160]. MIF has also been linked to processes involved in cell proliferation, differentiation, and tumor progression[161]. The mechanisms of the role MIF plays in tumor progression are unknown, but it is possible that increased MIF secretion by tumor cells aids in tumor promotion and survival by inducing the release of angiogenic factors[162, 163].

The loss or decrease in expression of both NADH-ubiquinone oxidoreductase and ubiquinol cytochrome c reductase has been identified in clear cell renal cell carcinoma. Both are present in mitochondria and take part in electron transport of the respiratory chain. The ubiquinol cytochrome c reductase is a component of the ubiquinol cytochrome c reductase complex. Its function is to generate an electrochemical potential coupled to ATP synthesis to transfer electrons from ubiquinol to cytochrome c. It is localized on the inner membranes of mitochondria. NADH-ubiquinone oxidoreductase catalyses the transfer of electrons from NADH to ubiquinone[101, 105].

Cytochrome c and the cytochrome c oxidase polypeptides are also involved in electron transport, which is consistent with findings that there is a severe decrease in

oxidative phosphorylation in clear cell renal cell carcinomas. The loss of activity of these proteins correlate with a severe loss of mitochondrial activity, which seems to favor a faster growth or increased invasiveness of these tumors[164]. These electron transport proteins were indicative of aberrant molecular presence outside of the histological borders [Figure 35].

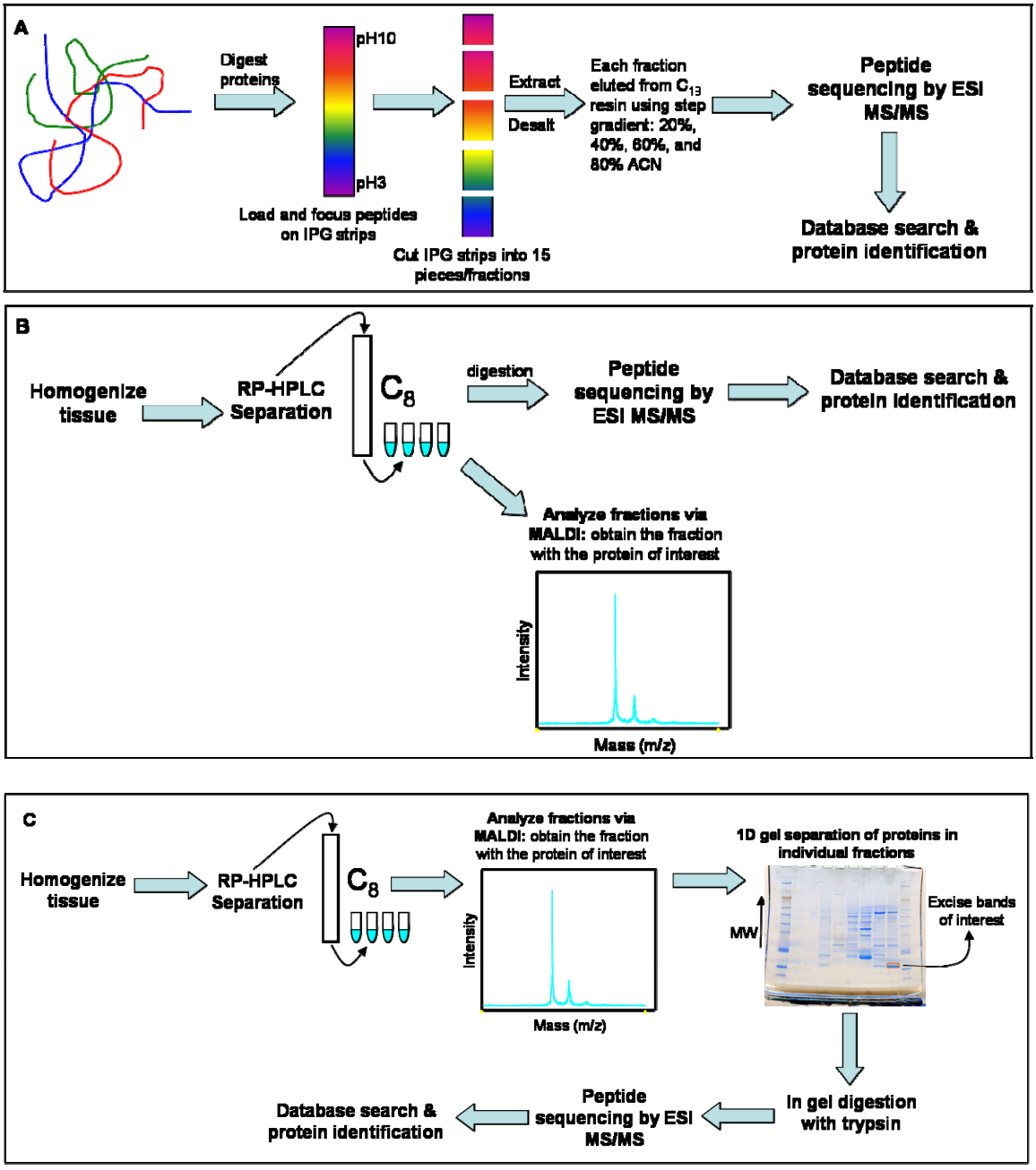


Figure 34: Schematic of the three methods used for protein identification.

Table 9: Summary of proteins identified. An asterisk (*) indicates a unique peptide. Letters A, B, and C in the "Methods" column indicates the protein identification method used as described by its corresponding letter in Figure 34.

MW	Protein Name	Accession #	% Coverage	Peptide Residue #'s	Method	Comment
11072	S100AA/Calpactin I	P60903	18	47-53*, 37-46*	C	Up in Tumor
10090	S100-A6/Calcyclin	P06703	24	41-47, 48-55, 56-62*	C	Down in Tumor
8577	COX6C	P09669	31	18-38	C	Down in Tumor
9368	NADH-ubiquinone oxidoreductase MLRQ subunit	O00483	53	36-47, 48-55, 56-63, 64-72, 77-81	C	Down in Tumor
11651	S100-A11/Calgizzarin	P31949	16	37-52*	A	Up in Tumor
10738	Ubiquinol-cytochrome c reductase complex 11 kDa protein	P07919	21	61-78	A	Down in Tumor
8714	Calbindin/Vitamin D-dependent Calcium BP	O75552	5	35-47	A	Down in Tumor
12272	Cytochrome C	P99999	57	28-38*, 39-52, 40-52, 56-73, 80-86, 92-99	B	Down in Tumor
10611	COX5B	P10606	24	50-56, 57-68, 58-68, 75-86*	B	Down in Tumor
6720	COX7A2	P14406	28	34-46*, 47-56*	B	Down in Tumor
5354	COX7C	P15954	14	26-34*	B	Down in Tumor
4890	COX82	P10176	13	26-34*	B	Up in Tumor
12345	MIF†	P14174			Other	Up in Tumor
4933	Thymosin β-10†	P63313			Other	Up in Tumor

† : Proteins were previously identified by intact fragmentation methods

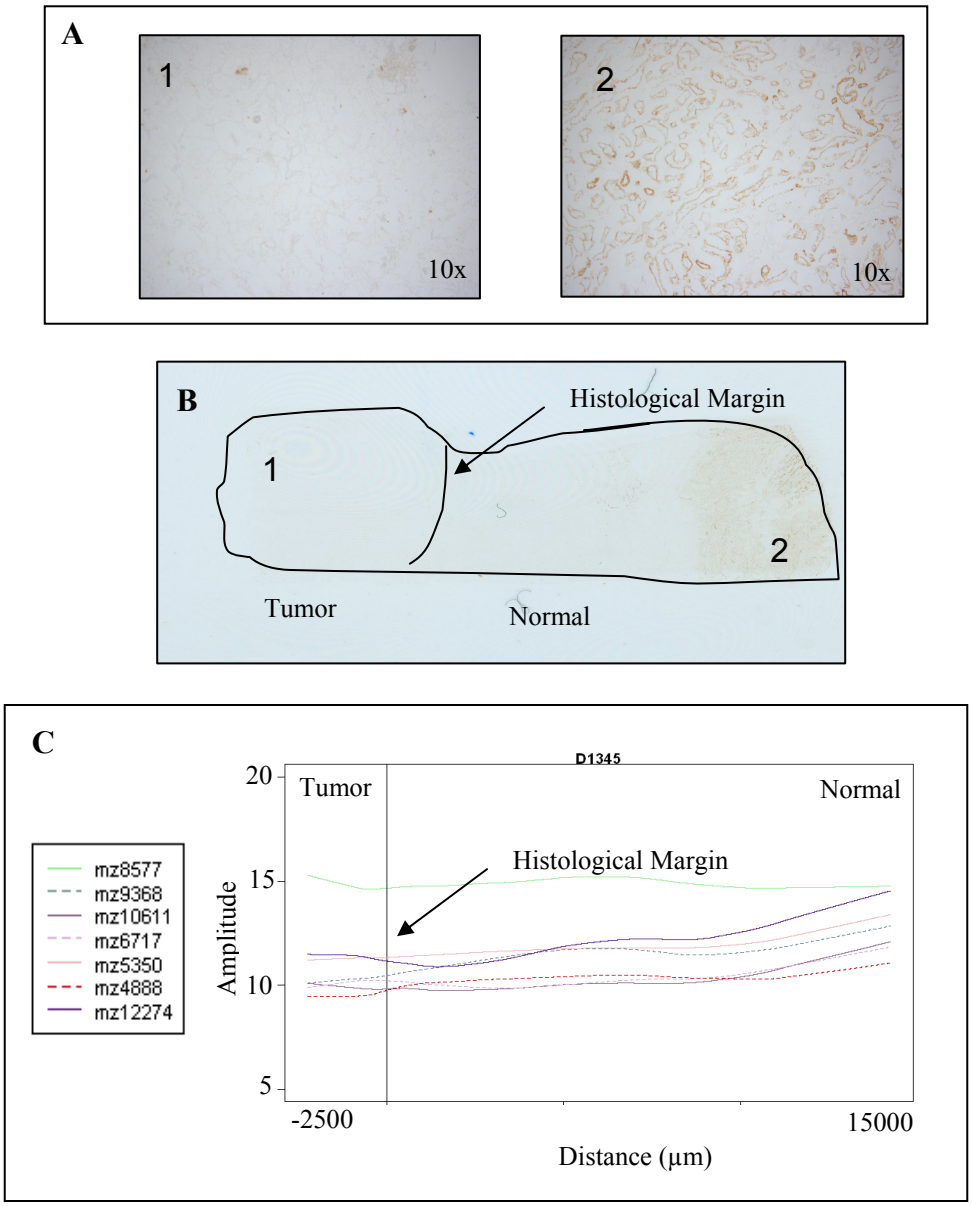


Figure 35: Cytochrome oxidase activity assay confirms localization of features. A.) Microscope images of tumor (1) and normal (2) regions of the tissue. B.) Scanned image of stained section shows localization. C.) LOWESS trend of features involved in mitochondrial electron transport.

Materials and Methods

To identify proteins of interest, tissue was homogenized with TPER (tissue protein extraction reagent) buffer (Pierce Biotechnology, Rockford, IL) according to the manufacturer's protocol. Protein supernatant was fractionated by RP-HPLC on a C_8 column (Grace Vydac, Hesperia, CA). One-dimensional SDS-PAGE separation of proteins was carried out on 10-20% Tricine gels (Invitrogen Corporation, Carlsbad, CA). In-gel and in-solution digests were carried out with trypsin (Promega Corporation, Madison, WI) according to the manufacturer's protocol. Peptide solutions were analyzed on a Deca XP Plus ion trap mass spectrometer and/or an LTQ linear ion trap mass spectrometer (Thermo Scientific, Waltham, MA), both equipped with a RP-C_{18} column.

For histochemistry, 4g sucrose was mixed with 90 µl of 0.1M phosphate buffer solution at pH 7.4. Then the solution was mixed with 50 mg of DAB/25 mg Cytochrome C type III (Sigma-Aldrich Co., C26506). After mixing, the solution was added to the slides and shaken for 45 minutes. Slides were removed and rinsed in a series of 3 x PBS (phosphate buffered saline), 3 x H2O, 3 x 70% EtOH, 3 x 95% EtOH, and 3 x xylene for 5 minutes each.

CHAPTER V

DISCUSSION AND CONCLUSIONS

Technological Perspectives

Application of MALDI MS to Clinical Pathology

The applicability of MALDI MS to the rapid evaluation of tumor margins has been proposed[165] and the supportive work presented here is the first, in-depth molecular assessment of tumor margin microenvironments by MALDI-TOF mass spectrometry. The rate of tumor recurrence post resection suggests that there are underlying molecular changes that go undetected in conventional diagnostic methods. To determine and better understand the molecular changes in and around the tumor margin requires the application of molecular technologies. Profiling and imaging MALDI MS technology has been applied to multiple diseased tissues and, with biocomputational tools, resulted in the identification of disease-state and patient prognostic-specific protein patterns[65, 67, 69-71, 125]. These studies suggest that proteomic information will become increasingly important in assessing disease progression, prognosis, and response to therapy.

With its ability to rapidly analyze thin tissue sections and provide for the visualization of hundreds of proteins simultaneously, it is believed that profiling and imaging MALDI MS technology will one day be used in pathology and diagnostic labs to enhance the quality of information provided to clinicians to improve upon conventional histopathological diagnoses[165, 166]. Conventional approaches rely on histochemistry and

immunohistochemistry procedures, which represent a challenge in the clinical setting. Histochemistry, such as hematoxylin and eosin staining, only allows for morphological characterization. Immunohistochemistry procedures, however, are difficult to standardize, and must be performed separately for each antibody; antibodies are not readily available and there lacks a standard, thorough procedure for assessing their cross-reactivity. Microarray analysis of antibody cross-reactivity produces worrisome results, including numerous cross-reactivities with both monoclonal and polyclonal antibodies, as well as preferential binding to non-specific proteins[167]. MALDI MS technology allows for the visualization of hundreds to thousands of individual proteins in the molecular weight range from 2000 to 200,000. Its application to direct tissue analysis allows for the direct correlation of individual protein signals with their distributions within specific regions of tissue. It is ideal for discovery studies in that it does not require prior knowledge of protein composition or require molecular-specific reagents.

This study has expanded on the applicability of profiling and imaging MALDI MS by demonstrating its ability to monitor patterns of molecular change around histological tumor margins to help facilitate the discovery of molecular signatures that will accommodate current histological procedures in defining the surgical margin status. To date, there have been some attempts to examine tumor margin status via molecular methods, but no currently available method demonstrates sufficient molecular sensitivity or specificity to challenge the superiority of surgical pathology[165]. In this study, MALDI MS was used to simultaneously measure hundreds of individual protein signatures, while retaining spatial information, in the microenvironments around the histologically-defined tumor margin and resulted in the identity and location of abnormal molecular

characteristics in the histologically normal tissue. This research shows that profiling intact tissue sections by MALDI MS, combined with biostatistics, offers a highly specific process of analyzing tumor margins to compliment the current clinical diagnostic tools in ensuring successful tumor extirpation.

The application of MALDI MS to the analysis of intact tissue sections was first reported nearly a decade ago[53], and though its potential superiority to IHC and electron microscopy as a molecular diagnostic tool has been demonstrated, several areas of improvement are needed before it can be of routine use in a clinical pathology lab[165, 166]. To be of value in a diagnostic laboratory, the complete assessment process must be rapid, or within a few hours of surgical resection. Although the data collection time has recently improved with increased laser repetition rates and enhanced electronics, critical advances are needed in terms of data processing. Spectral processing post acquisition involves both preprocessing and statistical computations. Currently, the processing time for individual samples can take days to complete, yet future software packages, allowing for a more automated and high throughput workflow will reduce the processing time to less than an hour. Analysis resolution, which is crucial to obtain cellular-specific as opposed to microenvironment specific information, is an ongoing improvement. Achievable resolution depends on the laser spot size on the target and matrix deposition. Some lasers can be focused down to 5 or 10 μm, but result in a sensitivity loss. Future advances in optics or sample matrices, such as nanoparticle additives, may enhance sensitivity at smaller laser spot sizes. With regards to matrix deposition, spray coating the matrix onto tissue allows for limitless resolution, but the most sensitive and reproducible data results from robotic deposition technologies, which are currently limited to approximately 200

μm in diameter. Continued improvements in the areas of optics, electronics, matrix deposition and biocomputations, as well as reductions in cost and instrument size, will push MALDI-TOF MS to the forefront of technologies capable of complementing histopathological diagnoses.

Importance of Tissue Collection and Preservation

Proper and uniform tissue collection and preservation protocols are vital to the successful application of proteomic technologies such as direct tissue profiling by MALDI MS to clinical diagnosis. Proteomic studies aiming to discover molecular events that preclude or directly result in phenotypic changes in tissue specimens must utilize samples in which little to no proteolysis, in excess of what occurs *in vivo*, has taken place. The process of degradation is slowed by low temperature, thus keeping the tissue specimen at temperatures slightly above freezing and ensuring rapid handling prior to snap-freezing in liquid nitrogen is essential to maintain protein integrity. The ability to transplant organs shows that tissue can be maintained *ex vivo* in the absence of blood circulation, but longevity differs between tissue types. The exact time that intact tissues can remain outside of the body without significant alterations to the proteome, however, is still unknown. There is a possibility that metabolites and macromolecules begin degrading as soon as the blood supply is no longer available. Metabolite levels can change within seconds or minutes, whereas macromolecule alterations occur anywhere from minutes to days[168, 169].

In a surgical setting, some time lapse between tissue resection and freezing is inevitable, but a better understanding of protein degradation in tissue is needed to design

and implement appropriate, standard sample handling protocols. Current standards of sample handling vary between repositories. As experienced in this study, many institutions that collect clinical specimens for research are not designed to achieve a time interval of less than two hours, indicating a need for tissue collection facilities to have uniform standard operating procedures. Another observation in this study was the lack of standard documentation procedures pertaining to the time between surgical resection and freezing. Uniform procedures and accurate documentation are essential to minimizing variability for proteomic studies. It is most likely that no routine protocols will be changed until there is solid evidence in support of a change, and currently, there remains to be a collective effort from proteome researchers to address this issue.

A second crucial aspect of tissue collection and preservation methods involves the conventional formalin fixation and paraffin embedding of resected tissue and the incompatibility of this method with proteomic analyses. Proteomic studies utilizing tissue require large amounts of sample, which is difficult to attain. There are vast amounts of archived tissue available, but the fixative, which covalently cross-links proteins, renders them difficult to analyze using routine proteomic protocols[169]. Formalin fixing and paraffin embedding tissue has become the gold standard for tissue storage and preservation of cellular morphology, but a new protocol is needed in order to accommodate both pathology labs and molecular profiling labs. Several solutions are proposed that may circumvent current and future problems with tissue preservation and storage. To potentially utilize the enormous amount of archived tissue, there needs to be a continued exploration into treatment methods that will allow the currently fixed and embedded archival tissue to be utilized in proteomic studies[169-172]. Secondly, a new

method of fixing and preservation is needed by which future samples can be preserved and archived. This could be done with new preservation strategies that are compatible with both microscopic and proteomic evaluation. Lastly, if these efforts prove to be of little value, then the collection and storage of clinical samples must consist of two parts, including one sample that is fixed and embedded for pathology use and one that is fresh-frozen for future proteomic use. Regardless, it is imperative that there be a uniform procedure followed by all collection facilities to minimize tissue and biological fluid variability and provide researchers with high quality samples.

Biological Perspectives

This study utilized MALDI MS to characterize the molecular microenvironment adjacent to histological tumor margins to determine if there are indications of aberrant molecular expression beyond the histological tumor border. Molecular features in the normal region adjacent to the histological tumor border that have an increased or decreased expression in correlation to their counterparts in the tumor could represent regions of aberrant cellular development. These cells could be morphologically indistinguishable from their normal complements, yet retain the potential for tumor invasion and metastasis.

Results of this study indicated that there were many proteins over-expressed and under-expressed in ccRCC tumors as compared to normal; however, those of particular interest are the proteins that indicated abnormal tissue in the normal region outside of the histologically defined tumor margin. Of the proteins identified, most are involved with mitochondrial electron transport. These proteins were consistently under-expressed in the

tumor as well as in the histologically normal tissue adjacent to the tumor. The decreased expression of proteins involved in electron transport has been observed in tumors of clear cell renal cell carcinoma[164, 173] but has not been observed in the immediate, adjacent normal tissue.

Mitochondrial Electron Transport

The electron transport chain, or respiratory system, is a series of membrane-associated complexes that mediate biochemical reactions to produce ATP. In short, the electron transport chain removes electrons from the donor, NADH, and passes them to a terminal electron acceptor, O_2 via a series of redox reactions. The reactions create a proton gradient across the mitochondrial inner membrane, resulting in the transmembrane proton gradient used to make ATP via ATP synthase in oxidative phosphorylation[174].

The process of electron transport consists of four complexes. Complex I, which consists of NADH dehydrogenase (synonym: NADH-ubiquinone oxidoreductase), removes two electrons from NADH and transfers them to the carrier, ubiquinone. Simultaneously, Complex I moves four protons across the membrane to produce a proton gradient later used to generate ATP through oxidative phosphorylation. Complex II, which consists of succinate dehydrogenase, funnels electrons into the quinone group by removing electrons from succinate and transferring them via FAD to the quinone. Complex III (cytochrome bc_1 complex) removes two electrons from the quinone and transfers them to two molecules of cytochrome *c*, a water-soluble electron transporter located on the outer surface of the membrane. Simultaneously, Complex III moves two protons across the membrane, producing a proton gradient consisting of four total

protons. When electron transfer is somehow hindered in Complexes I and III, there is a tendency for premature electron leakage to oxygen, resulting in the formation of superoxide. The last complex, Complex IV, consists of cytochrome *c* oxidase. It is responsible for removing four electrons from the four molecules of cytochrome *c* and transferring them to O_2, producing two molecules of water. Simultaneously, it also moves four protons across the membrane, producing a proton gradient. Finally, this proton gradient is utilized by the ATP synthase complex to make ATP via oxidative phosphorylation[174, 175].

Mitochondrial Electron Transport Proteins Identified in this Study

Proteins identified in this study that were under-expressed in tumor and involved in the electron transport system include cytochrome c, COX5B, COX6C, COX7A2, COX7C, COX8-2, a ubiquinol-cytochrome c reductase, and NADH-ubiquinone oxidoreductase (MLRQ subunit). Cytochrome c is an electron carrier protein. The oxidized form of the cytochrome c heme group accepts an electron from the heme group of the cytochrome c1 subunit of cytochrome reductase. Cytochrome c then transfers this electron to the cytochrome oxidase complex, which is the final step in the electron-transport process[176]. The cytochrome c oxidase subunits COX5B, COX6C, COX7A2, COX7C and COX8-2 are all subunits of Complex IV of the mitochondrial electron transport chain[177]. The function and exact location of the NADH-ubiquinone oxidoreductase (MLRQ subunit) protein remains unknown. It is believed to be a subunit associated with Complex I of the electron transport chain[178-181]. The ubiquinol-cytochrome c reductase is part of Complex III (cytochrome bc_1 complex). The proteins

identified represent a small fraction of the proteins found to be similarly expressed in the tumor and the immediate adjacent normal.

Mitochondrial Deficiency in Cancer

Mitochondrial deficiency in cancer cells was recognized almost eighty years ago by Otto Warburg[182], but only recently has it caught the attention of cancer researchers. The Warburg effect is described as the dependency of tumors on glycolysis rather than oxidative phosphorylation for ATP even in the presence of oxygen. The production of ATP through glycolysis is much more inefficient than its production through mitochondrial oxidative phosphorylation. It is perplexing, therefore, as to why cancer cells with this deficient energy metabolism, can persevere over other cells to develop a malignant population with drug-resistant potential.

Mechanisms of Mitochondrial Deficiency in ccRCC

The cause of mitochondrial electron transport deficiency in sporadic ccRCC remains unknown, but is believed to be related to mitochondrial DNA defects, mutations of nuclear DNA encoding mitochondrial proteins, mutations of proteins involved in their biogenesis, or mechanisms regulating protein translation or stability[164, 183]. Recent evidence suggests that the *VHL* tumor suppressor protein plays a role in stimulation of mitochondrial oxidative phosphorylation complex biogenesis[173, 184]. Clear cell renal cell carcinomas, however, are characterized by inactivation of this tumor suppressor gene. The VHL protein (pVHL), together with several other proteins (elongins C and B, Cul2 and Rbx1) that have E3 ubiquitin ligase activity, targets the alpha subunit of the

heterodimeric transcription factor HIF (hypoxia-inducible factor) for ubiquitination and subsequent destruction. In normoxic conditions, pVHL directs the ubiquitination of HIFα, which is then set for destruction. Under low oxygen conditions, a characteristic of tumors, the stabilized HIFα protein subunits accumulate in the nucleus where they form a complex with HIFβ. This complex is responsible for transcriptional activation of hypoxia adaptation factors. Consequences of this role include an increased synthesis of glucose transporters and glycolytic enzymes in order to maintain ATP production in the presence of low anaerobic metabolism compared to oxidative metabolism. In general, the hypoxia-inducible factors include genes involved in angiogenesis, erythropoiesis, glycolysis, cell growth and survival, and cell migration (CXCR4), suggesting that the VHL-HIF cascade of events are crucial for renal carcinoma progression[184].

Electron Transport Deficiency in the Adjacent Normal Kidney

Although the mitochondrial electron transport deficiency and related events in clear cell renal cell carcinoma have been examined, the mechanisms that trigger the events involved in tumor spread or invasion have been less studied. Particularly interesting is that results of this work suggest that this system is involved in tumor invasion into the adjacent normal tissue. Whether or not the mitochondrial deficiency is the primary or secondary cause of the cancer remains under scrutiny; however it now appears to be involved in tumor invasion into the adjacent normal tissue. Though there is insufficient research available to draw definitive conclusions, several mechanisms, including oxygen sensing processes[175, 185-187] and pH-mediated invasion[188, 189], may be involved that give rise to the observations in this study.

Although cells have a limited ability to generate ATP in the absence of oxygen, alternative mechanisms take over to sustain this supply when oxygen becomes limited. Most cells in the body are capable of sensing oxygen, but the oxygen sensor should initiate distinct signaling cascades when a slight deviation from normoxia is detected. There is a wide range of oxygen distribution in different tissues, suggesting that the threshold of activation may vary from tissue to tissue and cell to cell. The sensing mechanisms involved in cancer and how they relate to mitochondrial deficiency are unknown, yet a few pathways have been proposed, including prolyl hydroxylases and the mitochondria[175, 185].

The hypoxia inducible factor, responsible for activation of transcriptional responses to hypoxia, has been considered as oxygen-dependent; however, in the process of determining what activation mechanisms are responsible for its activity during hypoxia, it was discovered that the interaction between HIFα and pVHL is mediated by the hydroxylation of highly conserved proline residues within the ODD (oxygen-dependent degradation domain). The hydroxylation of these residues by prolyl hydroxylases facilitates interaction with the ubiquitin ligase, controlling protein stability. The regulation of prolyl hydroxylase activity by oxygen is therefore responsible for controlling the activation of HIF. The prolyl hydroxylases require 2-oxoglutarate and oxygen as substrates and a non-heme iron as a cofactor. When the activity of prolyl hydroxylases is blocked, there is an accumulation of HIFα followed by the activation of HIF-dependent gene expressions, as seen in ccRCC[175, 185, 186].

To be an oxygen sensor, the protein must be capable of initiating a response at the first sign of hypoxia. Research monitoring the sensitivity of prolyl hydroxylases to

oxygen levels by HIF stability observed an activation of this system at a very low oxygen level of 5%, with maximal activation at 0.5%, which is near anoxia[190]. This work was done in the *in vitro* environment, utilizing a small peptide from HIF rather than the full protein length. These factors may have introduced artificial differences in the enzyme activity, but finding a way to truly monitor this process in the intact cell in the *in vivo* environment is proving difficult[175]. Future work addressing more validated model systems will enable a better understanding of this pathway and how initial events involving the prolyl hydroxylases in oxygen sensing may contribute to tumor spread into the adjacent normal tissue.

Evidence has also suggested a mitochondrial role in oxygen sensing. As reviewed by Guzy and Shumacker[175] various studies have attempted to elucidate the potential pathways of mitochondrial oxygen sensing by evaluating the effect of electron transport chain complexes on HIF stabilization. In response to the evidence, the authors proposed that electron transport chain Complex III acts as an oxygen sensor by releasing ROS in response to hypoxia, which stabilizes HIF. During electron transport, a quasi-stable ubisemiquinone radical is repeatedly generated. Molecular oxygen, which is located within the membrane, is highly electrophilic and can potentially capture the electron from ubisemiquinone. The capture of an electron by oxygen yields superoxide, a reactive molecule that can potentially act in a signaling role or as a non-specific oxidizing agent that can contribute to cell damage. Since the cellular matrix is negative relative to the intermembrane space, it is proposed that the superoxide anion will then migrate to the intermembrane space, where it would be converted to hydrogen peroxide by superoxide dismutase. Hydrogen peroxide may contribute to tumor spread into the adjacent normal

tissue because of its capacity to cross membranes and target proteins for oxidation, resulting in functional consequences[175]. Evidence has also indicated that hydrogen peroxide may oxidize mitochondrial DNA polymerase γ, which could initiate further oxidative stress due to the loss of mitochondrial DNA replication and subsequent energy decline[191].

Though ROS invasion into the adjacent normal tissue is a probable event in decreased mitochondrial electron transport and enhanced tumor invasion, an additional mechanism, involving tumor acidosis, has been proposed that may likely be involved. Although each invasive cancer population appears to be the result of a unique genetic pathway traveled during carcinogenesis, most invasive cancers have in common an altered tumor metabolism, suggesting its role in favoring tumor invasion. Increased glycolysis in the tumor promotes acid excretion, which alters the microenvironment by substantially reducing intratumoral pH. This led Gatenby and Gawlinski[188] to propose that the H^+ ions in the tumor extracellular space diffuse along concentration gradients into adjacent normal tissues, which can cause cellular degradation or influence molecular events. An decrease in pH has been implicated in the reversible loss of VHL function by promoting its nucleolar sequestration, also in normoxic environments, with the subsequent stabilization of HIF[192]. Less research has been devoted to this potential pathway of invasion, but ccRCC invasion may be precluded by a combination of H+ and ROS diffusion into the adjacent normal tissue, resulting in HIF stabilization, retardation of the electron transport system, and increased glycolysis.

Factors Governing Distance of Compromise from Histological Margin

Factors governing how far cells are compromised outside the histological margin could be aspects of tumor hypoxia, aggressiveness, or a combination of these factors with the host immune system. Currently, there is no marker to define ccRCC tumor aggressiveness. In this study, tumor stage and grade were both utilized as an assessment of aggression in order to determine if the distance of compromised tissue outside of the histological margin could be correlated with aggressiveness; however the distance results could not be correlated with tumor stage or grade. This may indicate that the distance of compromise is independent of tumor aggressiveness. It could also be an indication of the deficiency in the current diagnostic system. The current histopathological inspection of the tumor is histological and may not best represent the aggressive nature of the tumor. There is a cascade of molecular events occurring in the tumor and its adjacent tissue that is not visualized with current histology procedures, thus, the tumor aggressiveness may best be defined by these underlying molecular events. Studies such as this may provide additional understanding of ccRCC aggression and invasion and may provide additional tools by which histologists can assess the tumor and its surrounding environment.

Status of Cells Expressing Aberrant Characteristics in Histologically Normal Tissue

Results from this project demonstrate that many intriguing yet complex events occur at tumor borders that contribute to the aberrant molecular changes occurring outside of the histological margin. Depending on the tissue and cancer type, these biological events may vary and may include one or more events. One hypothesis is the aberrant cells outside of the tumor margin are infiltrative tumor cells that go undetected

by conventional histology. Alternatively, these cells could be pre-cancerous cells that have already undergone malignant transformation at the molecular level yet show no phenotypic characteristics of tumor cells. A third hypothesis is that the aberrant cellular characteristics of these seemingly normal cells result from cross-talk or secretory interactions between the tumor and surrounding tissue. This signaling pathway could cause cells outside of the histological margin to express some features characteristic of the tumor, and possibly would no longer be expressed if the tumor were removed. There are an increasing number of investigators probing the biology of tumor spread, and perhaps one or more of the aforementioned hypotheses will one day explain the phenomenon of ccRCC tumor invasion and cancer recurrence.

Future Perspectives

Additional Insights into ccRCC Invasion

Despite recent discoveries in the tumorigenesis of ccRCC, much remains to be understood in what drives the invasion of this tumor. Though it is known that ROS is released from the electron transport chain, it is not understood how this may be induced by hypoxia and how this plays a role in oxygen sensing and tumor invasion[175]. An additional system of interest is the chemokine network. Previous studies have shown that chemokines play a major role regulating leukocyte trafficking and extravasation into sites of tissue inflammation. CXCR4, in particular has been demonstrated in cancer cells as mediating factor in tumor metastasis. Evidence suggests that CXCR4 may be regulated

by the HIF-1α subunit in ccRCC[193, 194]. Collectively, this information suggests that CXCR4 may play a crucial role in invasion into adjacent tissue, possibly through oxygen sensing or host inflammatory response mechanisms. Another interesting development is the role of membrane type-1 matrix metalloproteinase (MT1-MMP) as a key regulator of tumor progression through its functions as a matrix-degrading enzyme and as a cleaving factor to adhesion molecules and other MMPs. MT1-MMP has been identified as a potential transcriptional target of HIF-2alpha and has been indicated as a major mediator of tumor cell invasiveness in ccRCC[195]. Whether these events are triggered by oxygen sensing or other pathways is yet to be determined. Studies examining the presence or absence of these intriguing factors in the tissue immediately adjacent to the tumor would provide additional evidence of their role in tumor invasion. Future studies examining the gene expression levels in the tumor, histological margin, and adjacent normal and how they correlate to the nuclear encoded proteins identified in this study would provide insight into the potential genetic deficiencies of this cancer. Experiments measuring the levels of ROS, superoxide dismutase, and acidity of the tumor and adjacent tissue would provide additional evidence in support of or against the proposed ROS- and pH-mediated invasion mechanisms.

Assessing the Molecular Tumor Margins in Other Cancers

MALDI-MS methods for analyzing molecular changes in tumor margins has been successfully applied to ccRCC, with results indicating a glycolytic role in tumor invasion, either as a direct cause or a early indicator of compromised tissue in the normal microenvironment of the tumor. The molecular state of tumor margins is of high interest

and importance in all cancer types. Having varying degrees of aggression and potential to metastasize, different tumors may exhibit striking differences in their molecular signatures between the tumor, the histological margin, and the normally appearing tissue. Despite these differences, most tumors share the common characteristic of glycolytic dependency[188]. The high abundance of these features renders them easily detectable by mass spectrometry. The methods developed and implemented in this study may be easily applied to other cancer types. Results from these studies would provide new insights into understanding tumor margins in specific cancers as well as between cancers, while providing spatial molecular information that can be compared to conventional histochemistry approaches.

Three-Dimensional Imaging of Tumor Margins

This project examined molecular profiles of the tumor margin in a narrow, two-dimensional space. Results showed that tissue outside of the histological margin expressed certain tumor features. These aberrant characteristics were observed up to approximately 11 mm, depending on the sample, from the histological margin. It is not understood how this pattern exists in a three-dimensional distribution. The distance of compromise could be uniform in three dimensions or it could vary depending on tumor heterogeneity or the adjacent tissue. Experiments involving both invasive and non-invasive approaches to examining the three-dimensional structure of tumors have been performed, yet no molecular approach, such as mass spectrometry has been used to study the tumor volume. The three-dimensional visualization of protein mass spectra from MALDI MS in-situ analyses has been demonstrated[196] and current efforts are underway

to improve registration procedures and throughput as well as enhance image resolution to allow for substructure visualization. The ability to visualize molecular features of tumors in three dimensions would provide new biological insights into tumor invasion with the potential to improve the diagnosis of tumors and their classification in three dimensions in conjunction with current histochemistry approaches.

Early Detection of Renal Carcinoma

Although renal carcinoma is not one of the most prevalent diseases, it is one of the most silent killers, and its effects could be greatly minimized by early detection. Each year in the United States, there are approximately 30,000 new cases of renal cancer, which results in about 13,000 deaths per year. Unfortunately, this cancer, which is resistant to therapy, usually remains silent for many years before patients present themselves with symptoms due to metastatic disease. Kidney tumors can become quite large without causing any discomfort. Currently there are no reliable tests that can detect kidney cancer early. Routine urinalysis may find small amounts of blood in the urine, but many other causes of blood in the urine exist besides kidney cancer. Oftentimes, renal cancer does not present itself with blood in the urine, and when it does, the cancer is already quite large and has metastasized to other parts of the body[2, 76].

Proteomics has the power to determine potential early markers for renal carcinoma subtypes. Ultimately, biomarkers that may be traced from the primary tumor to biological fluids represent a noninvasive way to test for renal cancer. Since the kidneys serve as blood filterers and urine producers, urinalysis and serum analysis represent possible means for the early detection of ccRCC. Nonetheless, an early screening tool for

renal carcinoma would significantly reduce the amount of renal tumors found by chance. It is unclear whether or not glycolytic dependency in ccRCC is a result of nuclear- or mitochondrial-encoded defects, however, the depletion of these highly abundant electron transport proteins may indicate a very early event in tumor spread and may indicate an early event in initial, sporadic tumor development.

Conclusion

MALDI MS profiling of biological tissues has been demonstrated as a useful technology to monitor the patterns of molecular change traversing the histological tumor margin. The results of this analysis show that there are aberrant characteristics present outside of the histological margin. Depending on the sample, these changes were visible as far as 11000 micrometers after the histological tumor border. Identification of these changes indicated that increased glycolysis at the expense of mitochondrial oxidative phosphorylation plays a significant role in ccRCC tumor spread into the normal tissue. Assessment of molecular patterns in the tumor microenvironment indicated that there are very complex, yet distinctive ways in which molecules are expressed around the histological border. Given the vast problem with tumor recurrence, it is imperative to discover what drives these events. The mechanisms proposed to give rise to tumor invasion and mitochondrial deficiency in adjacent normal tissue include ROS- and pH-diffusion. Further elucidation of these events will not only provide additional insights into the molecular mechanisms of tumor invasion, but also facilitate the discovery of molecular markers to aid in current histological procedures that better define molecular tumor margins in an effort to reduce local recurrence. Identifying biological changes and

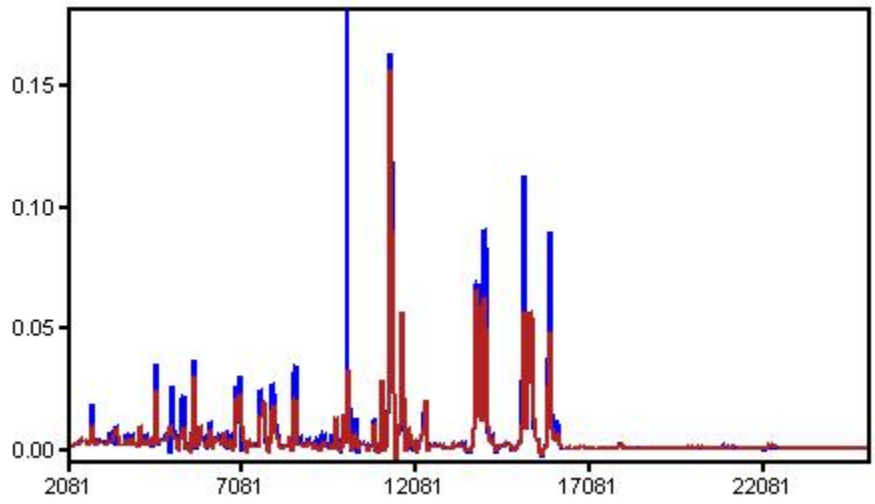
understanding the molecular events in and around tumor margins is vital to understanding therapeutic resistance and designing new therapeutic agents.

Appendix A

Comparison of Far Normal (FN) vs. Far Tumor (FT) from Tumor Grades III-IV

Data sets:
Selected spectra from six samples D1067, D529, D1340, D528, D1344, D1347

Plot of Group Averages: <group 6FN in blue, group 6FT in red>



Plot of Group Differences:

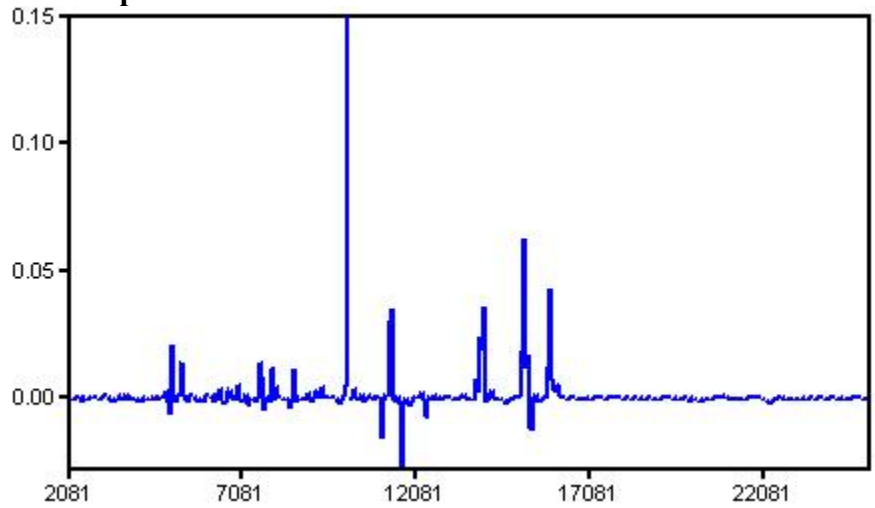


Table of features used in classifier and validation. (Selected Wil P < 0.0001):

MZ	ID	6FN Avg	6FT Avg	Wil P	T p	CV1	CV2
4112	25	.088	.375	2.520E-11	3.794E-06	.804	1.655
4385	27	.057	.244	1.255E-05	2.975E-06	.753	1.644
4887	33	.117	.051	1.863E-08	4.715E-10	.841	.687
4894	34	.111	.052	4.568E-07	8.163E-05	1.275	1.065
5042	38	.379	.193	3.507E-13	2.517E-12	.436	1.024
5061	39	.315	.099	4.284E-11	2.487E-05	1.591	1.32
5069	40	.56	.089	4.564E-09	7.833E-03	3.239	1.072
5260	42	.106	.067	1.959E-08	3.088E-09	.512	.526
5350	43	.572	.142	4.887E-11	2.142E-09	1.237	.532
5358	44	.425	.217	2.998E-09	1.008E-06	.907	.841
5668	47	.387	.305	2.507E-05	3.200E-05	.364	.463
5827	51	.105	.174	1.289E-06	9.425E-08	.568	.653
5932	52	.152	.061	3.131E-13	1.478E-09	.958	.443
6321	58	.141	.096	9.066E-06	5.181E-07	.581	.425
6329	59	.107	.072	3.856E-07	1.537E-06	.559	.594
6657	63	.178	.069	4.331E-09	2.958E-08	1.066	.67
6665	64	.155	.076	5.998E-08	3.944E-08	.863	.64
6716	65	.154	.087	1.640E-10	3.975E-12	.554	.444
6724	66	.128	.078	1.060E-06	5.448E-07	.678	.623
6952	70	.162	.085	3.800E-09	2.932E-09	.663	.834
7566	75	.594	.352	1.370E-05	6.639E-04	1.007	1.161
7680	77	.22	.302	2.729E-05	3.355E-06	.476	.477
8015	81	.235	.091	3.918E-08	1.491E-09	.974	.607
8024	82	.217	.114	4.881E-08	9.459E-09	.765	.59
8566	89	.704	.466	6.534E-07	8.777E-06	.56	.799
8586	90	.208	.13	2.755E-05	1.516E-05	.739	.736
8713	91	.125	.072	1.885E-05	7.219E-05	.908	.993
9191	93	.143	.073	3.490E-06	3.565E-05	1.09	.963
9248	95	.103	.055	5.653E-07	3.374E-05	.974	1.031
9366	97	.222	.068	1.224E-09	1.451E-09	1.127	.479
9374	98	.195	.075	9.662E-11	1.610E-09	.98	.611
9620	100	.165	.065	1.156E-13	2.324E-12	.816	.524
9835	104	.034	.089	3.178E-07	2.879E-10	.84	.901
9960	106	.214	.27	3.889E-05	1.417E-03	.607	.449
10090	107	3.213	1.929	3.376E-09	2.514E-06	.471	1.188
10132	108	.24	.152	1.611E-11	4.181E-06	.491	1.022
10299	111	.287	.19	3.640E-08	1.246E-04	.586	1.027
10468	112	.081	.064	8.540E-06	1.114E-02	.481	.875
11040	114	.117	.263	1.651E-06	3.791E-07	.963	1.012
11072	115	.333	.909	3.017E-06	3.536E-08	.894	1.098
11101	116	.076	.341	3.952E-09	4.154E-10	.908	1.207
11609	121	.178	.307	1.183E-05	5.288E-07	.703	.736
11653	122	.839	1.618	1.857E-06	6.412E-08	.771	.794
11732	123	.065	.127	1.728E-05	1.338E-07	.579	.877
11967	125	.071	.158	1.115E-05	7.505E-07	.895	1.041
12764	130	.05	.095	4.423E-05	3.885E-06	.685	.984
13425	134	.102	.068	1.447E-07	4.900E-08	.52	.476
13904	140	.56	.201	5.107E-15	0.000E+00	.641	.977
14093	144	.114	.254	3.327E-07	2.577E-10	.684	.802
14623	148	.07	.121	7.806E-05	1.179E-05	.846	.829
15171	152	.477	.345	2.985E-06	3.432E-02	.839	1.454
15276	153	.974	.622	2.442E-15	0.000E+00	.289	.462
15379	155	.347	.57	5.336E-09	3.199E-10	.473	.543

15423	156	.083	.179	9.465E-10	8.153E-10	.801	.784
19924	170	.06	.099	2.604E-05	8.407E-02	3.163	1.361
20139	171	.028	.058	9.478E-05	7.494E-06	.897	1.084
20198	172	.034	.179	1.182E-06	8.037E-06	.723	1.832
20927	173	.104	.053	4.778E-07	4.130E-08	.826	.676
21267	175	.017	.027	5.655E-05	7.125E-06	.84	.674
22202	177	.05	.472	2.389E-12	1.867E-06	.855	1.884
22268	178	.03	.081	3.791E-06	2.914E-08	.569	1.117

K table : PDiff: 0; 6FN- group size: 107, 6FT- group size: 107

k	Errors	Misses	G1 Errors	G1 Misses	G2 Errors	G2 Misses
7	9	0	6	0	3	0
11	9	0	4	0	5	0
5	9	0	5	0	4	0
1	9	0	6	0	3	0
13	10	0	5	0	5	0
3	10	0	6	0	4	0
9	10	0	6	0	4	0
15	13	0	6	0	7	0
17	14	0	5	0	9	0
19	19	0	7	0	12	0

Complete feature table

Used as Classifier	MZ	ID	6FN Avg	6FT Avg	Wil P	T p	CV1	CV2
True	15276	153	.974	.622	2.442E-15	0.000E+00	.289	.462
True	13904	140	.56	.201	5.107E-15	0.000E+00	.641	.977
True	9620	100	.165	.065	1.156E-13	2.324E-12	.816	.524
True	5932	52	.152	.061	3.131E-13	1.478E-09	.958	.443
True	5042	38	.379	.193	3.507E-13	2.517E-12	.436	1.024
True	22202	177	.05	.472	2.389E-12	1.867E-06	.855	1.884
True	10132	108	.24	.152	1.611E-11	4.181E-06	.491	1.022
True	4112	25	.088	.375	2.520E-11	3.794E-06	.804	1.655
True	5061	39	.315	.099	4.284E-11	2.487E-05	1.591	1.32
True	5350	43	.572	.142	4.887E-11	2.142E-09	1.237	.532
True	9374	98	.195	.075	9.662E-11	1.610E-09	.98	.611
True	6716	65	.154	.087	1.640E-10	3.975E-12	.554	.444
True	15423	156	.083	.179	9.465E-10	8.153E-10	.801	.784
True	9366	97	.222	.068	1.224E-09	1.451E-09	1.127	.479
True	5358	44	.425	.217	2.998E-09	1.008E-06	.907	.841
True	10090	107	3.213	1.929	3.376E-09	2.514E-06	.471	1.188
True	6952	70	.162	.085	3.800E-09	2.932E-09	.663	.834
True	11101	116	.076	.341	3.952E-09	4.154E-10	.908	1.207
True	6657	63	.178	.069	4.331E-09	2.958E-08	1.066	.67
True	5069	40	.56	.089	4.564E-09	7.833E-03	3.239	1.072
True	15379	155	.347	.57	5.336E-09	3.199E-10	.473	.543

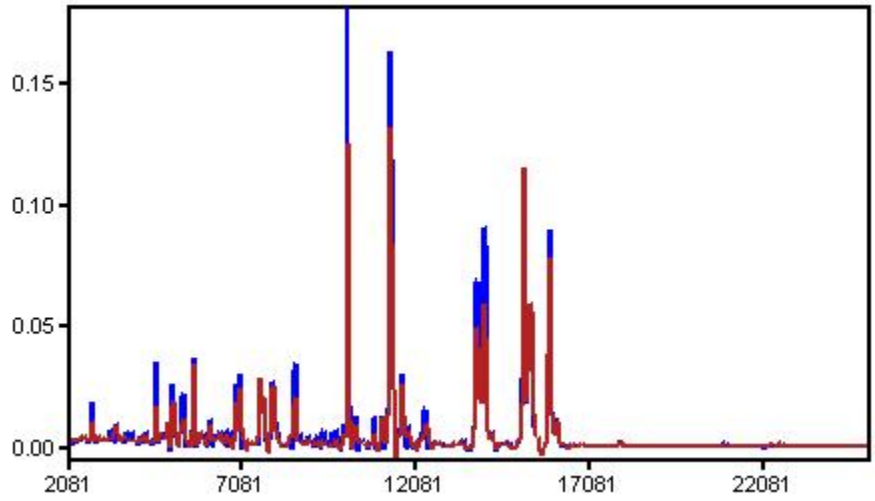
True	4887	33	.117	.051	1.863E-08	4.715E-10	.841	.687
True	5260	42	.106	.067	1.959E-08	3.088E-09	.512	.526
True	10299	111	.287	.19	3.640E-08	1.246E-04	.586	1.027
True	8015	81	.235	.091	3.918E-08	1.491E-09	.974	.607
True	8024	82	.217	.114	4.881E-08	9.459E-09	.765	.59
True	6665	64	.155	.076	5.998E-08	3.944E-08	.863	.64
True	13425	134	.102	.068	1.447E-07	4.900E-08	.52	.476
True	9835	104	.034	.089	3.178E-07	2.879E-10	.84	.901
True	14093	144	.114	.254	3.327E-07	2.577E-10	.684	.802
True	6329	59	.107	.072	3.856E-07	1.537E-06	.559	.594
True	4894	34	.111	.052	4.568E-07	8.163E-05	1.275	1.065
True	20927	173	.104	.053	4.778E-07	4.130E-08	.826	.676
True	9248	95	.103	.055	5.653E-07	3.374E-05	.974	1.031
True	8566	89	.704	.466	6.534E-07	8.777E-06	.56	.799
True	6724	66	.128	.078	1.060E-06	5.448E-07	.678	.623
True	20198	172	.034	.179	1.182E-06	8.037E-06	.723	1.832
True	5827	51	.105	.174	1.289E-06	9.425E-08	.568	.653
True	11040	114	.117	.263	1.651E-06	3.791E-07	.963	1.012
True	11653	122	.839	1.618	1.857E-06	6.412E-08	.771	.794
True	15171	152	.477	.345	2.985E-06	3.432E-02	.839	1.454
True	11072	115	.333	.909	3.017E-06	3.536E-08	.894	1.098
True	9191	93	.143	.073	3.490E-06	3.565E-05	1.09	.963
True	22268	178	.03	.081	3.791E-06	2.914E-08	.569	1.117
True	10468	112	.081	.064	8.540E-06	1.114E-02	.481	.875
True	6321	58	.141	.096	9.066E-06	5.181E-07	.581	.425
True	11967	125	.071	.158	1.115E-05	7.505E-07	.895	1.041
True	11609	121	.178	.307	1.183E-05	5.288E-07	.703	.736
True	4385	27	.057	.244	1.255E-05	2.975E-06	.753	1.644
True	7566	75	.594	.352	1.370E-05	6.639E-04	1.007	1.161
True	11732	123	.065	.127	1.728E-05	1.338E-07	.579	.877
True	8713	91	.125	.072	1.885E-05	7.219E-05	.908	.993
True	5668	47	.387	.305	2.507E-05	3.200E-05	.364	.463
True	19924	170	.06	.099	2.604E-05	8.407E-02	3.163	1.361
True	7680	77	.22	.302	2.729E-05	3.355E-06	.476	.477
True	8586	90	.208	.13	2.755E-05	1.516E-05	.739	.736
True	9960	106	.214	.27	3.889E-05	1.417E-03	.607	.449
True	12764	130	.05	.095	4.423E-05	3.885E-06	.685	.984
True	21267	175	.017	.027	5.655E-05	7.125E-06	.84	.674
True	14623	148	.07	.121	7.806E-05	1.179E-05	.846	.829
True	20139	171	.028	.058	9.478E-05	7.494E-06	.897	1.084
False	8032	83	.156	.114	1.007E-04	6.586E-04	.619	.687
False	10268	109	.247	.174	1.109E-04	7.418E-05	.622	.617
False	9954	105	.19	.24	1.306E-04	3.307E-03	.666	.503
False	8961	92	.127	.093	1.486E-04	1.521E-04	.574	.618
False	9237	94	.102	.075	1.511E-04	1.631E-02	.851	.994
False	12277	127	.466	.255	1.537E-04	6.629E-06	.923	.768
False	15128	151	3.243	2.562	1.645E-04	8.970E-02	.85	1.2
False	13382	133	.077	.063	1.687E-04	5.069E-04	.381	.468
False	9750	102	.226	.334	2.153E-04	3.842E-03	1.067	.889
False	9744	101	.211	.281	2.896E-04	3.300E-02	1.109	.852
False	14229	146	.159	.123	2.920E-04	1.594E-02	.585	.986
False	7558	74	.536	.403	3.434E-04	5.376E-02	.928	1.262
False	5651	46	.762	.589	3.937E-04	5.686E-04	.48	.605
False	4282	26	.108	.076	3.968E-04	3.018E-04	.647	.742
False	4736	30	.048	.13	4.196E-04	1.582E-05	1.103	1.417
False	4566	29	.652	.473	4.229E-04	3.510E-02	.941	1.311
False	22421	179	.029	.09	4.542E-04	3.188E-05	.844	1.615

False	8406	86	.074	.111	5.441E-04	5.590E-03	.867	1.078
False	6431	60	.146	.116	5.441E-04	1.400E-01	.81	1.482
False	5529	45	.068	.117	6.603E-04	3.505E-06	.551	.848
False	8448	87	.215	.239	1.126E-03	6.406E-01	1.821	1.425
False	8091	85	.108	.076	1.285E-03	2.034E-03	.773	.838
False	9777	103	.117	.083	1.443E-03	8.630E-04	.764	.683
False	15924	160	.412	.426	1.463E-03	8.366E-01	.86	1.377
False	12346	128	.382	.566	1.550E-03	8.384E-04	.935	.768
False	8038	84	.137	.106	1.959E-03	9.793E-03	.688	.731
False	14217	145	.171	.151	2.619E-03	2.139E-01	.489	.938
False	11390	120	.644	.536	3.544E-03	1.432E-02	.481	.613
False	3484	18	.085	.265	3.836E-03	1.668E-02	3.066	2.735
False	8455	88	.187	.212	3.938E-03	5.554E-01	1.783	1.353
False	14701	150	.084	.082	4.016E-03	8.888E-01	1.655	.552
False	6120	54	.137	.112	4.232E-03	2.460E-02	.661	.593
False	5170	41	.07	.154	4.457E-03	2.040E-03	2.062	1.549
False	7929	78	.554	.501	4.815E-03	4.317E-01	.851	1.023
False	13711	135	.139	.197	5.035E-03	8.256E-04	.694	.759
False	15870	159	2.512	2.444	5.133E-03	8.545E-01	.911	1.253
False	3433	14	.133	.263	5.298E-03	8.599E-04	1.027	1.426
False	10281	110	.24	.175	5.538E-03	1.309E-03	.733	.632
False	14011	142	2.031	1.763	5.538E-03	1.543E-01	.564	.889
False	15333	154	1.132	1.287	5.861E-03	1.687E-02	.328	.43
False	7937	79	.575	.434	6.278E-03	2.489E-02	.9	.876
False	13153	131	.13	.297	6.476E-03	6.512E-03	1.806	1.958
False	3325	9	.169	.184	7.325E-03	7.371E-01	1.299	2.209
False	7006	71	.549	.48	7.415E-03	2.005E-01	.609	.928
False	11349	119	2.363	2.037	7.645E-03	4.139E-02	.422	.642
False	16077	162	.349	.38	8.678E-03	6.097E-01	.956	1.386
False	4917	35	.078	.097	8.835E-03	2.213E-02	.69	.707
False	4747	31	.114	.487	9.049E-03	4.078E-05	1.989	1.834
False	3421	12	.124	.084	1.000E-02	5.809E-02	1.453	1.385
False	5819	50	.149	.233	1.164E-02	1.952E-03	.853	1.051
False	5675	48	.381	.324	1.366E-02	6.416E-02	.559	.709
False	7655	76	.346	.313	1.381E-02	2.171E-02	.298	.352
False	19685	169	.024	.036	1.381E-02	1.891E-03	.659	.96
False	3273	7	.135	.132	1.821E-02	8.941E-01	.892	1.558
False	21770	176	.034	.031	1.964E-02	2.575E-01	.592	.804
False	2748	3	.458	.519	2.326E-02	6.155E-01	1.41	2.106
False	16483	163	.05	.057	2.554E-02	3.829E-01	.697	1.329
False	4559	28	.919	.59	2.607E-02	3.364E-02	1.534	1.257
False	11308	118	3.665	3.361	2.716E-02	1.812E-01	.337	.591
False	15852	158	1.74	2.201	2.772E-02	1.547E-01	.946	1.321
False	21129	174	.029	.065	2.989E-02	7.574E-06	.548	1.225
False	23417	181	.018	.022	3.080E-02	2.087E-02	.566	.613
False	13784	138	2.042	1.864	4.677E-02	3.574E-01	.621	.825
False	16033	161	.207	.271	5.056E-02	1.279E-01	.95	1.433
False	3478	17	.082	.173	5.242E-02	1.613E-03	1.026	1.639
False	2933	4	.085	.103	5.434E-02	3.237E-01	.889	1.63
False	13360	132	.07	.067	5.887E-02	6.690E-01	.53	.85
False	5696	49	.167	.142	6.070E-02	8.707E-02	.744	.611
False	6893	68	.478	.437	6.097E-02	3.824E-01	.655	.837
False	11265	117	.497	.469	6.340E-02	5.691E-01	.628	.826
False	6311	57	.082	.072	7.056E-02	2.702E-01	.753	.878
False	6548	62	.112	.139	7.237E-02	4.409E-02	.731	.808
False	3577	19	.042	.055	7.768E-02	1.051E-02	.688	.801
False	14053	143	.398	.392	7.800E-02	8.825E-01	.666	.889

False	6541	61	.097	.124	8.499E-02	1.985E-02	.588	.848
False	4767	32	.106	.072	8.742E-02	1.212E-02	1.187	.835
False	6112	53	.203	.157	8.812E-02	2.100E-02	.876	.65
False	16785	164	.054	.067	1.091E-01	2.518E-02	.651	.773
False	4964	37	.78	1.688	1.163E-01	4.091E-03	2.082	1.659
False	17940	166	.055	.059	1.172E-01	4.823E-01	.789	.71
False	14658	149	.07	.068	1.198E-01	8.587E-01	1.41	.552
False	3895	22	.158	.094	1.207E-01	2.859E-02	1.818	1.008
False	18418	167	.032	.035	1.243E-01	1.086E-01	.417	.465
False	12229	126	.339	.261	1.340E-01	1.921E-01	1.552	1.235
False	17894	165	.094	.095	1.379E-01	9.550E-01	.972	.906
False	6166	55	.062	.076	1.448E-01	1.131E-01	.893	.961
False	3427	13	.198	.205	1.509E-01	8.032E-01	.958	1.303
False	7953	80	.129	.127	1.520E-01	9.080E-01	.951	1.098
False	3901	23	.124	.122	1.726E-01	8.699E-01	.874	1.207
False	9521	99	.069	.044	1.882E-01	1.211E-02	1.331	1.049
False	12689	129	.074	.12	1.900E-01	3.673E-02	1.072	1.748
False	3319	8	.084	.075	1.948E-01	4.344E-01	1.066	1.094
False	11818	124	.272	.301	1.967E-01	4.805E-01	1.068	1.011
False	2477	0	.053	.085	1.991E-01	9.859E-03	1.014	1.362
False	13726	136	.25	.313	2.016E-01	3.108E-02	.628	.828
False	3362	10	.114	.234	2.035E-01	2.679E-03	1.201	1.645
False	2955	5	.059	.088	2.091E-01	2.127E-02	.973	1.315
False	3439	15	.137	.331	2.098E-01	3.358E-02	2.207	2.689
False	13981	141	.507	.569	2.149E-01	2.455E-01	.523	.858
False	23860	182	.02	.021	2.280E-01	6.205E-01	.422	.653
False	3722	21	.154	.277	2.307E-01	1.658E-02	1.435	1.722
False	13818	139	.779	.711	2.499E-01	5.287E-01	1.089	.988
False	13761	137	1.163	1.341	2.570E-01	1.385E-01	.437	.844
False	6910	69	.188	.203	2.577E-01	5.368E-01	.929	.857
False	9338	96	.068	.113	2.584E-01	4.999E-03	.734	1.373
False	3716	20	.111	.135	2.627E-01	3.135E-01	1.343	1.432
False	15822	157	.361	.622	2.649E-01	8.769E-03	1.17	1.493
False	3267	6	.185	.083	2.693E-01	7.959E-03	2.087	.925
False	2554	2	.05	.07	2.722E-01	3.203E-02	.983	1.202
False	7432	73	.146	.14	2.848E-01	9.038E-01	2.406	2.16
False	10837	113	.386	.536	3.289E-01	5.892E-02	1.182	1.266
False	18718	168	.019	.018	3.467E-01	6.483E-01	.608	.609
False	2512	1	.063	.127	3.475E-01	5.890E-03	1.368	1.747
False	23121	180	.018	.019	3.723E-01	5.573E-01	.559	.688
False	3368	11	.13	.434	3.993E-01	1.809E-02	3.065	2.897
False	7028	72	.132	.143	4.010E-01	3.797E-01	.622	.728
False	24714	183	.02	.02	4.019E-01	8.829E-01	.498	.664
False	4042	24	.083	.139	4.234E-01	7.243E-02	1.281	2.164
False	6226	56	.066	.078	4.355E-01	2.989E-01	1.172	1.178
False	6879	67	.229	.251	4.433E-01	2.306E-01	.412	.668
False	3472	16	.084	.093	4.529E-01	4.509E-01	.865	1.133
False	4937	36	.186	.383	4.687E-01	1.013E-03	1.247	1.473
False	14490	147	.075	.078	4.960E-01	8.196E-01	1.618	1.237

Comparison of Far Normal (FN) vs. Near Margin Normal (NN) in Tumor Grades III and IV

Plot of Group Averages: <group 6FN in blue, group 6NN in red>



Plot of Group Differences:

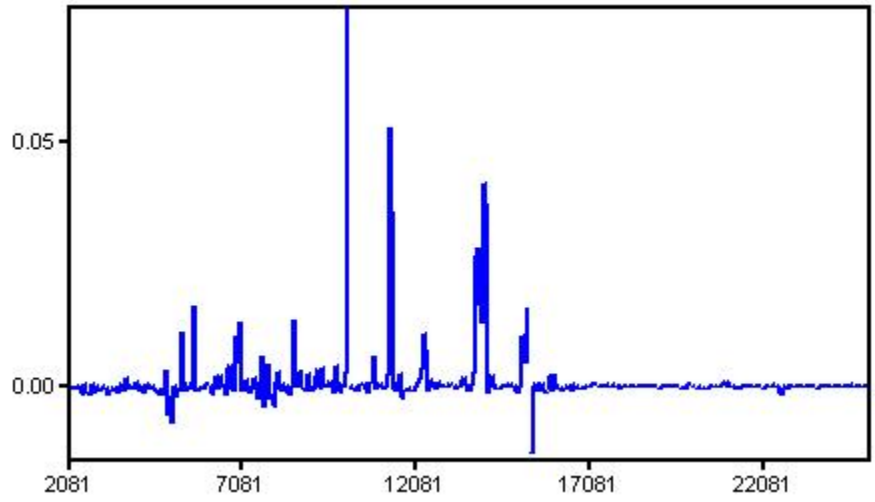


Table of features used in classifier and validation. (Selected Wil P < 0.0005):

MZ	ID	6FN Avg	6NN Avg	Wil P	T p	CV1	CV2
3325	12	.168	.066	5.240E-06	1.148E-05	1.298	1.318
3427	15	.197	.09	2.280E-08	9.845E-07	.963	1.194
3580	19	.046	.068	4.035E-05	2.572E-05	.568	.665
3901	22	.113	.066	7.186E-06	2.319E-05	.819	.98
4566	27	.652	.48	8.055E-05	3.659E-02	.941	1.207
4888	31	.12	.046	4.539E-11	3.707E-12	.816	.695
4917	32	.081	.11	1.985E-04	9.945E-04	.671	.632
5054	37	.197	.713	9.237E-07	4.629E-05	.777	1.786
5069	39	.56	.47	3.888E-05	6.288E-01	3.239	1.307
5351	42	.598	.173	3.606E-10	5.144E-09	1.188	.631
5357	43	.493	.159	0.000E+00	1.200E-12	.915	.365
5651	44	.771	.537	5.469E-07	3.199E-07	.483	.498
5932	49	.152	.075	4.537E-06	3.859E-07	.958	.476
6324	53	.141	.093	1.671E-07	1.156E-07	.547	.501
6657	57	.18	.081	3.536E-06	2.929E-07	1.038	.545
6664	58	.163	.072	3.761E-10	1.846E-09	.877	.546
6716	59	.155	.079	8.759E-14	2.775E-14	.556	.537
6724	60	.128	.058	5.884E-15	6.368E-13	.677	.631
7644	70	.352	.233	2.331E-15	2.220E-16	.274	.418
8017	75	.246	.106	4.243E-08	4.809E-09	.916	.655
8091	77	.108	.053	5.503E-09	1.486E-08	.773	.872
8176	78	.026	.078	1.040E-07	6.594E-09	1.239	1.074
8404	79	.075	.089	1.187E-04	6.098E-02	.859	.54
8560	81	.533	.323	8.534E-10	5.576E-08	.55	.774
8571	82	.689	.49	1.752E-07	4.680E-03	.632	1.168
8584	83	.229	.147	5.469E-07	1.770E-04	.724	1.015
8713	84	.124	.057	1.415E-06	4.741E-08	.909	.833
8958	85	.128	.071	3.613E-09	6.762E-12	.563	.479
9191	86	.137	.095	4.308E-06	4.677E-02	1.046	1.751
9366	87	.222	.072	6.353E-09	5.226E-09	1.127	.587
9618	89	.167	.06	1.754E-14	3.326E-13	.83	.468
9744	90	.212	.106	5.034E-05	3.683E-05	1.112	1.036
9750	91	.23	.117	3.677E-05	4.080E-05	1.097	.994
9775	92	.124	.056	4.428E-12	2.482E-11	.762	.506
9960	93	.215	.126	7.436E-10	2.144E-09	.604	.539
10096	97	3.043	2.14	4.885E-07	2.986E-07	.414	.574
10836	101	.398	.287	1.512E-04	1.311E-01	1.188	2.081
11308	106	3.638	2.842	4.679E-06	2.873E-05	.338	.519
12275	114	.476	.178	1.171E-10	1.021E-09	.946	.947
12346	115	.382	.271	1.689E-04	6.725E-03	.932	.828
13381	120	.077	.059	1.384E-07	3.587E-05	.373	.572
13425	121	.102	.059	4.227E-12	1.373E-12	.52	.451
13437	122	.107	.057	5.496E-13	3.022E-12	.604	.426
13904	126	.561	.383	5.080E-05	1.507E-04	.643	.816
15261	136	.777	.547	6.921E-07	5.755E-07	.347	.682
15379	139	.347	.504	2.360E-05	2.554E-05	.473	.673
20920	155	.107	.061	1.871E-04	1.529E-06	.856	.539
22517	160	.054	.096	6.271E-11	1.109E-02	2.509	1.014

K table : PDiff: 0; 6FN- group size: 107, 6NN- group size: 106

k	Errors	Misses	G1 Errors	G1 Misses	G2 Errors	G2 Misses
1	19	0	11	0	8	0
3	20	0	14	0	6	0
5	24	0	15	0	9	0
9	27	0	20	0	7	0
7	27	0	19	0	8	0
15	30	0	19	0	11	0
11	30	0	19	0	11	0
13	31	0	21	0	10	0
17	33	0	20	0	13	0
19	38	0	22	0	16	0

Complete feature table

Used in Classifier	MZ	ID	6FN Avg	6NN Avg	Wil P	T p	CV1	CV2
True	5357	43	.493	.159	0.000E+00	1.200E-12	.915	.365
True	7644	70	.352	.233	2.331E-15	2.220E-16	.274	.418
True	6724	60	.128	.058	5.884E-15	6.368E-13	.677	.631
True	9618	89	.167	.06	1.754E-14	3.326E-13	.83	.468
True	6716	59	.155	.079	8.759E-14	2.775E-14	.556	.537
True	13437	122	.107	.057	5.496E-13	3.022E-12	.604	.426
True	13425	121	.102	.059	4.227E-12	1.373E-12	.52	.451
True	9775	92	.124	.056	4.428E-12	2.482E-11	.762	.506
True	4888	31	.12	.046	4.539E-11	3.707E-12	.816	.695
True	22517	160	.054	.096	6.271E-11	1.109E-02	2.509	1.014
True	12275	114	.476	.178	1.171E-10	1.021E-09	.946	.947
True	5351	42	.598	.173	3.606E-10	5.144E-09	1.188	.631
True	6664	58	.163	.072	3.761E-10	1.846E-09	.877	.546
True	9960	93	.215	.126	7.436E-10	2.144E-09	.604	.539
True	8560	81	.533	.323	8.534E-10	5.576E-08	.55	.774
True	8958	85	.128	.071	3.613E-09	6.762E-12	.563	.479
True	8091	77	.108	.053	5.503E-09	1.486E-08	.773	.872
True	9366	87	.222	.072	6.353E-09	5.226E-09	1.127	.587
True	3427	15	.197	.09	2.280E-08	9.845E-07	.963	1.194
True	8017	75	.246	.106	4.243E-08	4.809E-09	.916	.655
True	8176	78	.026	.078	1.040E-07	6.594E-09	1.239	1.074
True	13381	120	.077	.059	1.384E-07	3.587E-05	.373	.572
True	6324	53	.141	.093	1.671E-07	1.156E-07	.547	.501
True	8571	82	.689	.49	1.752E-07	4.680E-03	.632	1.168
True	10096	97	3.043	2.14	4.885E-07	2.986E-07	.414	.574
True	8584	83	.229	.147	5.469E-07	1.770E-04	.724	1.015
True	5651	44	.771	.537	5.469E-07	3.199E-07	.483	.498
True	15261	136	.777	.547	6.921E-07	5.755E-07	.347	.682
True	5054	37	.197	.713	9.237E-07	4.629E-05	.777	1.786
True	8713	84	.124	.057	1.415E-06	4.741E-08	.909	.833
True	6657	57	.18	.081	3.536E-06	2.929E-07	1.038	.545
True	9191	86	.137	.095	4.308E-06	4.677E-02	1.046	1.751
True	5932	49	.152	.075	4.537E-06	3.859E-07	.958	.476
True	11308	106	3.638	2.842	4.679E-06	2.873E-05	.338	.519

True	3325	12	.168	.066	5.240E-06	1.148E-05	1.298	1.318
True	3901	22	.113	.066	7.186E-06	2.319E-05	.819	.98
True	15379	139	.347	.504	2.360E-05	2.554E-05	.473	.673
True	9750	91	.23	.117	3.677E-05	4.080E-05	1.097	.994
True	5069	39	.56	.47	3.888E-05	6.288E-01	3.239	1.307
True	3580	19	.046	.068	4.035E-05	2.572E-05	.568	.665
True	9744	90	.212	.106	5.034E-05	3.683E-05	1.112	1.036
True	13904	126	.561	.383	5.080E-05	1.507E-04	.643	.816
True	4566	27	.652	.48	8.055E-05	3.659E-02	.941	1.207
True	8404	79	.075	.089	1.187E-04	6.098E-02	.859	.54
True	10836	101	.398	.287	1.512E-04	1.311E-01	1.188	2.081
True	12346	115	.382	.271	1.689E-04	6.725E-03	.932	.828
True	20920	155	.107	.061	1.871E-04	1.529E-06	.856	.539
True	4917	32	.081	.11	1.985E-04	9.945E-04	.671	.632
False	9522	88	.068	.028	5.577E-04	2.649E-05	1.352	1.118
False	5062	38	.308	.741	5.937E-04	4.761E-04	1.612	1.562
False	7432	66	.146	.079	6.933E-04	8.306E-02	2.396	2.392
False	19746	153	.026	.035	7.205E-04	1.603E-02	.969	.76
False	10088	96	3.032	2.286	7.316E-04	1.152E-03	.55	.714
False	10045	94	.173	.107	8.331E-04	3.300E-04	.922	.901
False	12649	116	.062	.068	8.394E-04	1.661E-01	.644	.415
False	15289	137	1.064	.846	9.055E-04	7.496E-05	.272	.565
False	7684	72	.174	.228	9.617E-04	7.274E-04	.535	.592
False	13823	125	.603	.456	1.059E-03	9.365E-02	1.132	1.3
False	4931	33	.132	.212	1.067E-03	2.296E-05	.702	.782
False	13781	124	1.985	1.669	1.100E-03	8.664E-02	.585	.895
False	5668	45	.405	.34	1.141E-03	3.968E-03	.393	.487
False	14630	132	.072	.076	1.275E-03	5.715E-01	.878	.478
False	10271	99	.25	.18	1.293E-03	1.121E-04	.626	.542
False	10131	98	.252	.201	1.303E-03	1.341E-03	.494	.52
False	5042	36	.39	.338	1.352E-03	7.142E-02	.45	.715
False	11350	107	2.344	2.024	1.740E-03	4.165E-02	.431	.622
False	3421	14	.124	.052	1.991E-03	1.114E-04	1.453	1.108
False	6626	56	.039	.108	2.490E-03	4.259E-05	1.496	1.473
False	11390	108	.656	.551	2.559E-03	1.450E-02	.474	.574
False	10297	100	.292	.224	3.077E-03	2.345E-03	.61	.618
False	2272	0	.039	.043	3.205E-03	4.828E-01	1.455	.875
False	8033	76	.154	.125	3.688E-03	2.053E-02	.619	.713
False	2477	1	.051	.116	3.737E-03	5.438E-05	1.019	1.32
False	14012	127	2.068	1.739	4.126E-03	7.289E-02	.578	.841
False	6113	50	.204	.157	4.610E-03	3.116E-02	.851	.91
False	4560	26	.955	.884	4.610E-03	7.174E-01	1.511	1.626
False	22705	161	.028	.036	4.640E-03	6.595E-03	.644	.694
False	3273	10	.135	.09	7.969E-03	1.226E-03	.897	.834
False	4937	34	.188	.3	7.969E-03	5.312E-03	1.251	1.126
False	3439	17	.14	.251	8.367E-03	2.711E-02	2.121	1.672
False	22199	159	.051	.038	9.896E-03	1.731E-02	.865	.867
False	21268	156	.017	.027	1.007E-02	5.898E-04	.845	.991
False	3895	21	.157	.07	1.100E-02	2.770E-03	1.833	1.007
False	12226	113	.335	.402	1.126E-02	4.300E-01	1.561	1.767
False	3365	13	.137	.196	1.301E-02	1.127E-01	1.957	1.424
False	19267	152	.017	.022	1.516E-02	4.340E-03	.465	.611
False	6953	63	.158	.128	1.639E-02	4.827E-02	.694	.845
False	14054	128	.391	.346	1.639E-02	2.498E-01	.666	.889
False	3319	11	.085	.05	1.781E-02	8.883E-04	1.103	.986
False	10076	95	1.259	.986	1.860E-02	1.316E-01	1.213	1.074
False	7027	65	.138	.118	2.061E-02	9.632E-02	.595	.808

False	3478	18	.081	.126	2.280E-02	2.674E-02	1.141	1.49
False	6891	61	.463	.413	2.764E-02	2.401E-01	.601	.811
False	2747	4	.444	.307	2.952E-02	8.273E-02	1.433	1.648
False	14222	130	.174	.169	3.280E-02	7.492E-01	.519	.789
False	15332	138	1.133	1.246	3.379E-02	1.187E-01	.327	.518
False	6312	52	.088	.081	3.463E-02	4.766E-01	.742	1.111
False	21539	157	.018	.021	3.673E-02	3.006E-02	.529	.569
False	7006	64	.549	.493	4.182E-02	2.633E-01	.596	.819
False	11268	105	.538	.449	4.489E-02	3.157E-02	.607	.591
False	4767	30	.105	.104	6.086E-02	9.595E-01	1.176	.875
False	6911	62	.173	.147	6.140E-02	2.424E-01	.951	1.092
False	23369	162	.017	.021	6.248E-02	3.711E-02	.695	.78
False	3722	20	.153	.092	6.386E-02	1.084E-02	1.426	1.249
False	16790	147	.055	.06	6.470E-02	2.389E-01	.643	.56
False	3434	16	.134	.189	6.640E-02	7.184E-02	1.121	1.449
False	11818	112	.275	.344	6.669E-02	2.524E-01	1.08	1.572
False	11609	109	.173	.156	8.574E-02	2.835E-01	.68	.748
False	2553	3	.05	.057	8.679E-02	3.239E-01	.983	.955
False	5696	47	.167	.143	1.049E-01	9.274E-02	.744	.515
False	11678	111	.171	.212	1.073E-01	6.657E-02	.797	.877
False	13720	123	.202	.195	1.106E-01	7.259E-01	.655	.839
False	16488	146	.051	.064	1.110E-01	9.603E-02	.729	1.177
False	11259	104	.401	.341	1.162E-01	7.976E-02	.677	.644
False	2957	8	.058	.07	1.184E-01	2.957E-01	.972	1.459
False	5675	46	.374	.351	1.210E-01	4.738E-01	.589	.724
False	14092	129	.112	.143	1.219E-01	2.531E-02	.682	.819
False	12684	117	.075	.074	1.260E-01	9.326E-01	1.082	1.922
False	2929	6	.062	.067	1.302E-01	5.060E-01	.877	.789
False	4115	24	.087	.078	1.331E-01	2.129E-01	.6	.702
False	6548	55	.112	.101	1.345E-01	3.262E-01	.731	.847
False	2935	7	.086	.114	1.474E-01	1.066E-01	1.164	1.308
False	11652	110	.841	.796	1.500E-01	6.255E-01	.776	.87
False	5170	40	.07	.225	1.511E-01	1.307E-04	2.068	1.701
False	8448	80	.215	.146	1.601E-01	1.100E-01	1.821	1.446
False	14696	133	.089	.066	1.661E-01	1.775E-01	1.924	.596
False	7930	73	.585	.623	1.678E-01	6.142E-01	.839	.974
False	4282	25	.109	.116	1.769E-01	6.135E-01	.636	1.038
False	4747	29	.114	.15	1.786E-01	2.788E-01	1.984	1.708
False	4673	28	.068	.093	1.809E-01	4.281E-02	.769	1.268
False	2783	5	.064	.048	1.857E-01	1.079E-01	1.491	.864
False	15128	134	3.24	3.417	1.880E-01	6.728E-01	.857	.968
False	13150	119	.132	.144	1.904E-01	8.084E-01	1.814	3.109
False	7937	74	.575	.548	1.965E-01	6.990E-01	.9	.889
False	15826	140	.424	.615	1.990E-01	3.236E-02	1.127	1.277
False	20762	154	.029	.029	2.053E-01	9.458E-01	.437	.553
False	4041	23	.081	.085	2.254E-01	8.351E-01	1.309	1.77
False	21774	158	.033	.039	2.288E-01	1.719E-01	.626	1.02
False	12763	118	.05	.044	2.376E-01	1.990E-01	.695	.686
False	15172	135	.473	.597	2.383E-01	1.021E-01	.857	1.116
False	15866	141	2.517	2.867	2.432E-01	3.363E-01	.909	1.035
False	11038	102	.109	.095	2.502E-01	3.122E-01	.95	.971
False	6431	54	.147	.146	2.516E-01	9.584E-01	.806	.937
False	11073	103	.328	.321	2.545E-01	8.575E-01	.869	.803
False	18423	150	.031	.03	2.830E-01	8.778E-01	.377	.466
False	5263	41	.112	.141	2.846E-01	1.662E-02	.43	.816
False	4964	35	.78	2.322	2.952E-01	1.447E-04	2.082	1.63
False	15913	142	.518	.686	3.123E-01	5.426E-02	.888	1.122

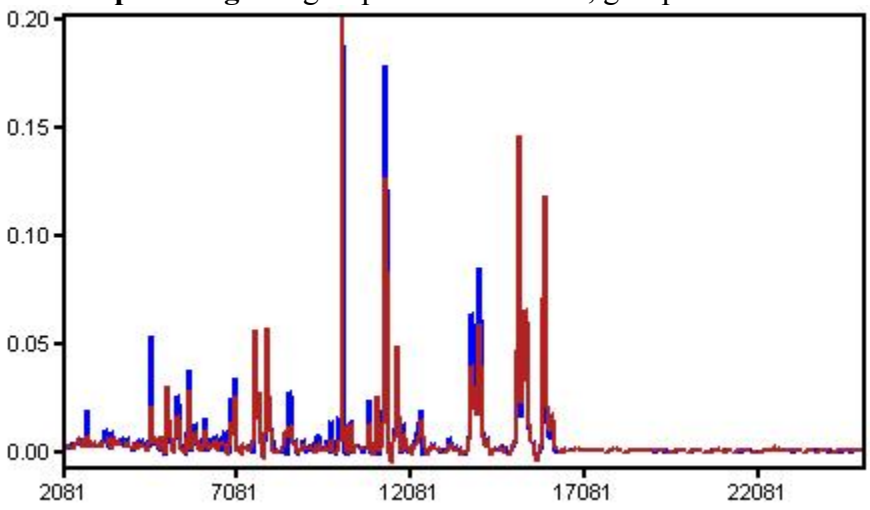
False	24664	164	.019	.021	3.147E-01	2.663E-01	.524	.611
False	6226	51	.065	.076	3.147E-01	3.523E-01	1.188	1.152
False	18792	151	.018	.021	3.266E-01	1.614E-01	.545	.68
False	7580	69	.133	.14	3.395E-01	6.994E-01	.882	1.074
False	15929	143	.369	.497	3.404E-01	4.045E-02	.854	1.126
False	17924	149	.074	.063	3.560E-01	1.381E-01	.884	.594
False	7566	68	.594	.585	3.576E-01	9.187E-01	1.007	.95
False	5820	48	.151	.139	3.676E-01	4.734E-01	.837	.87
False	7559	67	.57	.647	3.794E-01	3.491E-01	.924	1.014
False	16077	145	.343	.447	4.033E-01	6.565E-02	.949	1.066
False	2512	2	.065	.06	4.033E-01	6.503E-01	1.295	1.256
False	14491	131	.075	.11	4.051E-01	1.130E-01	1.604	1.745
False	23831	163	.021	.022	4.128E-01	4.647E-01	.502	.646
False	17878	148	.095	.077	4.224E-01	9.767E-02	1.029	.702
False	7666	71	.377	.391	4.224E-01	5.329E-01	.343	.469
False	16032	144	.21	.297	4.346E-01	2.089E-02	.958	1.1
False	3267	9	.185	.081	4.372E-01	7.958E-03	2.086	1.187

Appendix B

Comparison of Far Normal (FN) vs. Near Margin Normal (NN) of Cortex Border Tissue (TC) from Tumor Grades III-IV

Data sets:
Near Margin Normal (NN) -- selected spectra from samples D1067, D529, D1340
Far Normal (FN) -- selected spectra from samples D1067, D529, D1340

Plot of Group Averages: <group 3TC FN in blue, group 3TC NN in red>



Plot of Group Differences:

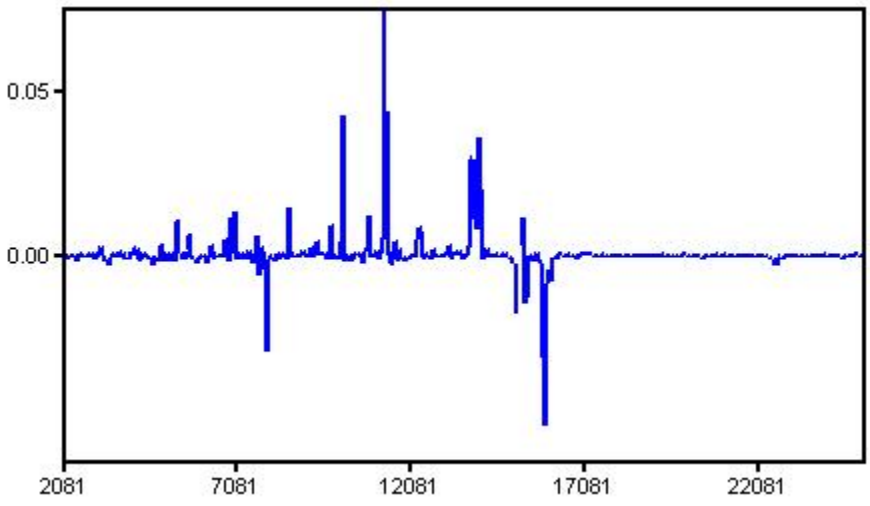


Table of features used in classifier and validation
(Selected features with Wil P < 0.00002)

MZ	ID	3TC FN Avg	3TC NN Avg	Wil P	T p	CV1	CV2
3427	17	.177	.061	2.471E-06	4.776E-06	.935	.949
4888	39	.111	.051	1.226E-08	1.778E-08	.589	.573
5351	52	.437	.205	1.533E-05	1.847E-05	.824	.529
5357	53	.34	.174	4.718E-10	1.486E-09	.508	.346
5647	56	.598	.409	1.127E-06	9.381E-07	.31	.461
6320	66	.132	.085	1.519E-08	6.641E-06	.394	.589
6716	72	.165	.088	1.370E-07	4.595E-08	.498	.535
6724	73	.117	.059	3.310E-07	7.382E-08	.561	.552
7643	84	.349	.244	4.444E-06	8.221E-07	.295	.425
7761	87	.2	.074	1.533E-05	7.748E-04	1.244	1.177
8560	98	.469	.24	1.619E-05	2.036E-05	.697	.757
9184	104	.186	.045	4.989E-06	8.926E-05	1.339	.835
9367	106	.179	.077	2.399E-06	3.619E-05	.931	.571
9616	107	.131	.067	9.199E-09	2.280E-08	.549	.416
9775	110	.114	.061	3.763E-07	1.233E-07	.526	.497
9953	111	.244	.116	5.948E-09	1.546E-08	.548	.625
9958	112	.269	.137	3.298E-09	2.477E-08	.522	.55
11305	125	3.757	2.609	1.619E-05	5.925E-05	.355	.575
13425	142	.083	.05	7.539E-07	6.481E-07	.458	.488
22506	184	.037	.118	2.546E-06	2.786E-05	1.326	1.079

K table : PDiff: 0; 3TC FN- group size: 54, 3TC NN- group size: 53

k	Errors	Misses	G1 Errors	G1 Misses	G2 Errors	G2 Misses
3	15	0	11	0	4	0
5	17	0	12	0	5	0
1	18	0	10	0	8	0
7	20	0	15	0	5	0
11	21	0	14	0	7	0
15	22	0	16	0	6	0
17	22	0	15	0	7	0
19	23	0	16	0	7	0
13	23	0	15	0	8	0
9	24	0	16	0	8	0

Complete feature table

Used as Classifier	MZ	ID	3TC FN Avg	3TC NN Avg	Wil P	T p	CV1	CV2
True	5357	53	.34	.174	4.718E-10	1.486E-09	.508	.346
True	9958	112	.269	.137	3.298E-09	2.477E-08	.522	.55
True	9953	111	.244	.116	5.948E-09	1.546E-08	.548	.625
True	9616	107	.131	.067	9.199E-09	2.280E-08	.549	.416
True	4888	39	.111	.051	1.226E-08	1.778E-08	.589	.573
True	6320	66	.132	.085	1.519E-08	6.641E-06	.394	.589
True	6716	72	.165	.088	1.370E-07	4.595E-08	.498	.535
True	6724	73	.117	.059	3.310E-07	7.382E-08	.561	.552
True	9775	110	.114	.061	3.763E-07	1.233E-07	.526	.497
True	13425	142	.083	.05	7.539E-07	6.481E-07	.458	.488
True	5647	56	.598	.409	1.127E-06	9.381E-07	.31	.461
True	9367	106	.179	.077	2.399E-06	3.619E-05	.931	.571
True	3427	17	.177	.061	2.471E-06	4.776E-06	.935	.949
True	22506	184	.037	.118	2.546E-06	2.786E-05	1.326	1.079
True	7643	84	.349	.244	4.444E-06	8.221E-07	.295	.425
True	9184	104	.186	.045	4.989E-06	8.926E-05	1.339	.835
True	7761	87	.2	.074	1.533E-05	7.748E-04	1.244	1.177
True	5351	52	.437	.205	1.533E-05	1.847E-05	.824	.529
True	11305	125	3.757	2.609	1.619E-05	5.925E-05	.355	.575
True	8560	98	.469	.24	1.619E-05	2.036E-05	.697	.757
False	8568	99	.495	.281	2.011E-05	1.183E-05	.573	.659
False	5363	54	.151	.085	2.181E-05	2.269E-05	.634	.6
False	3325	13	.17	.046	2.629E-05	1.492E-04	1.298	1.414
False	6542	68	.111	.064	5.890E-05	4.756E-05	.578	.786
False	13380	141	.067	.055	7.972E-05	5.441E-02	.348	.682
False	3273	11	.17	.067	8.174E-05	2.836E-05	.944	.851
False	11387	128	.668	.501	2.205E-04	1.285E-03	.373	.547
False	10834	120	.626	.433	2.783E-04	1.674E-01	.948	1.898
False	13779	144	1.658	1.138	3.195E-04	4.557E-03	.523	.865
False	4566	33	.775	.539	5.846E-04	9.544E-02	.95	1.325
False	13820	145	.43	.29	6.379E-04	3.022E-02	.753	1.156
False	8713	101	.128	.076	7.264E-04	2.566E-04	.707	.598
False	11348	127	2.589	1.929	7.422E-04	4.163E-03	.429	.631
False	9750	109	.319	.155	8.260E-04	4.134E-04	.939	.876
False	6656	71	.127	.08	9.184E-04	9.511E-04	.723	.534
False	9744	108	.303	.146	1.230E-03	5.038E-04	.952	.932
False	7025	78	.148	.103	1.230E-03	2.924E-03	.513	.748
False	3580	21	.048	.071	1.881E-03	2.348E-03	.492	.691
False	6897	76	.337	.232	1.919E-03	1.561E-02	.683	.917
False	8580	100	.197	.139	2.077E-03	1.364E-03	.547	.505
False	8958	102	.129	.085	2.160E-03	1.602E-04	.575	.435
False	12271	134	.374	.225	2.382E-03	6.770E-03	.863	1.005
False	6890	75	.402	.299	2.428E-03	2.162E-02	.539	.793
False	11318	126	3.302	2.419	2.428E-03	1.566E-03	.426	.582
False	5668	58	.416	.33	2.833E-03	5.342E-03	.361	.489
False	5055	47	.275	1.085	2.943E-03	1.107E-03	.646	1.628
False	14049	148	.372	.286	3.115E-03	5.608E-02	.567	.866
False	5071	49	.166	.389	3.175E-03	8.886E-03	1.325	1.477
False	7861	88	.096	.142	3.175E-03	7.502E-03	.825	.671
False	8090	94	.098	.055	3.423E-03	4.958E-03	.994	.975
False	14009	147	1.939	1.497	4.436E-03	5.033E-02	.533	.848
False	3895	24	.251	.08	4.770E-03	1.945E-03	1.523	1.073

False	5654	57	.633	.479	5.127E-03	1.085E-02	.508	.611
False	3716	22	.154	.063	5.127E-03	1.371E-03	1.223	1.121
False	6112	64	.284	.216	5.913E-03	9.080E-02	.721	.958
False	15278	160	.936	.743	6.457E-03	9.191E-03	.359	.557
False	6876	74	.201	.164	7.293E-03	4.240E-02	.416	.612
False	5062	48	.321	.93	8.364E-03	6.075E-03	1.582	1.626
False	8015	92	.219	.134	8.949E-03	2.180E-03	.821	.594
False	13901	146	.426	.31	9.255E-03	2.008E-02	.653	.729
False	13720	143	.186	.153	1.110E-02	2.194E-01	.673	.977
False	6547	69	.111	.078	1.147E-02	2.249E-02	.674	.937
False	10259	117	.208	.149	1.204E-02	1.762E-02	.679	.729
False	7004	77	.515	.433	1.306E-02	1.656E-01	.5	.8
False	5695	60	.198	.14	1.327E-02	1.535E-02	.787	.525
False	4560	32	1.45	1.322	1.579E-02	7.237E-01	1.286	1.408
False	3319	12	.113	.058	1.629E-02	1.679E-03	1.012	.906
False	8448	97	.4	.207	1.760E-02	1.310E-02	1.215	1.323
False	12687	137	.115	.11	1.760E-02	8.829E-01	.902	1.856
False	7685	86	.189	.226	1.842E-02	7.202E-02	.556	.47
False	4384	31	.071	.052	1.842E-02	2.831E-02	.73	.687
False	7497	81	.083	.094	2.111E-02	3.467E-01	.838	.504
False	11266	124	.54	.412	2.111E-02	2.898E-02	.602	.657
False	12344	135	.472	.345	2.143E-02	1.029E-01	1.051	.781
False	2750	6	.386	.364	2.412E-02	8.569E-01	1.406	1.944
False	14622	154	.079	.087	2.748E-02	5.253E-01	.868	.606
False	5675	59	.381	.309	3.167E-02	9.604E-02	.594	.713
False	19920	178	.102	.023	3.167E-02	4.194E-02	2.738	.391
False	11678	130	.228	.315	3.348E-02	3.127E-02	.75	.755
False	13150	139	.239	.241	3.395E-02	9.804E-01	1.248	2.566
False	4041	25	.104	.102	3.998E-02	9.384E-01	1.329	2.007
False	14683	155	.132	.078	4.052E-02	1.605E-01	2.067	.598
False	3421	16	.163	.059	4.217E-02	2.724E-03	1.457	1.147
False	10095	115	3	2.504	4.274E-02	6.190E-02	.36	.635
False	5048	46	.357	.574	4.331E-02	9.742E-03	.672	.971
False	22198	183	.064	.042	4.388E-02	2.684E-02	.935	1.018
False	2477	1	.057	.115	4.505E-02	1.919E-02	1.101	1.447
False	14218	151	.164	.146	4.565E-02	3.593E-01	.551	.789
False	3721	23	.156	.084	4.747E-02	2.419E-02	1.3	1.309
False	14221	152	.159	.14	4.809E-02	3.281E-01	.558	.799
False	14091	149	.155	.141	5.063E-02	4.264E-01	.505	.748
False	14210	150	.161	.145	5.129E-02	4.054E-01	.529	.772
False	11258	123	.42	.332	5.261E-02	7.861E-02	.653	.701
False	6572	70	.077	.072	5.604E-02	7.572E-01	.779	1.535
False	11816	131	.366	.512	5.675E-02	1.795E-01	.996	1.371
False	12763	138	.065	.054	5.891E-02	1.544E-01	.613	.689
False	3267	10	.313	.096	6.039E-02	4.298E-03	1.678	1.317
False	7432	80	.239	.135	6.659E-02	1.620E-01	1.984	1.888
False	15379	162	.415	.489	6.740E-02	1.183E-01	.469	.583
False	10127	116	.249	.217	6.740E-02	1.551E-01	.421	.584
False	3363	14	.184	.174	6.822E-02	8.382E-01	1.041	1.79
False	12225	133	.501	.615	6.904E-02	4.601E-01	1.331	1.482
False	10274	118	.236	.187	6.904E-02	6.569E-02	.649	.627
False	11038	121	.161	.133	7.071E-02	2.131E-01	.751	.859
False	13357	140	.071	.064	7.241E-02	4.763E-01	.547	.922
False	6430	67	.125	.108	7.503E-02	2.393E-01	.58	.715
False	4107	26	.104	.066	7.503E-02	6.070E-02	1.259	.895
False	9072	103	.065	.047	7.681E-02	5.717E-02	.875	.869
False	8174	95	.045	.086	7.862E-02	2.195E-03	.657	1.061

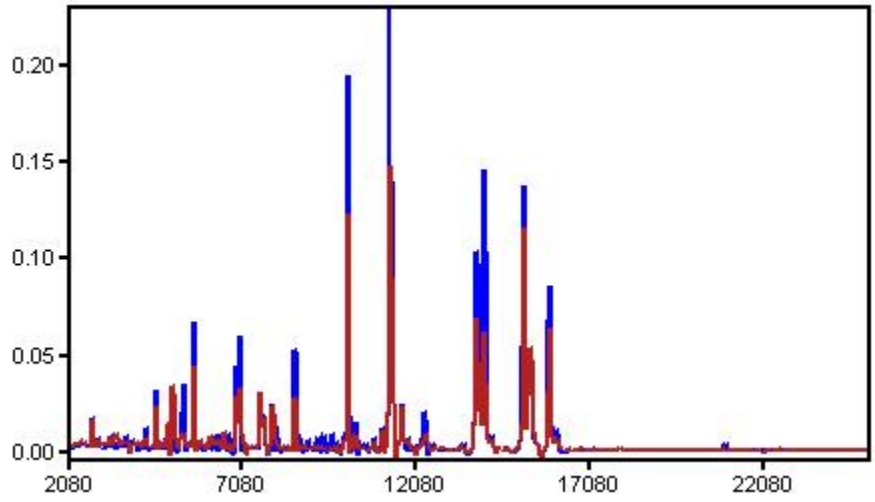
False	15330	161	1.247	1.338	8.140E-02	3.823E-01	.322	.479
False	3478	20	.13	.128	9.748E-02	9.583E-01	1.032	1.556
False	5263	51	.11	.165	1.040E-01	7.372E-03	.395	.853
False	4115	27	.092	.08	1.086E-01	3.205E-01	.658	.824
False	5819	61	.222	.189	1.086E-01	2.493E-01	.647	.791
False	21770	182	.024	.024	1.133E-01	8.700E-01	.549	.713
False	19757	177	.026	.032	1.194E-01	8.336E-02	.483	.673
False	2744	5	.637	.141	1.232E-01	3.349E-03	1.87	1.168
False	10088	114	3.278	2.757	1.245E-01	1.284E-01	.501	.68
False	20185	179	.041	.032	1.297E-01	1.240E-01	.924	.632
False	5932	63	.112	.082	1.297E-01	2.266E-02	.788	.45
False	7950	91	.147	.216	1.310E-01	4.083E-02	.867	.961
False	4253	29	.055	.09	1.405E-01	6.305E-02	1.006	1.35
False	12643	136	.063	.07	1.405E-01	3.152E-01	.543	.469
False	10069	113	.611	.442	1.490E-01	2.093E-01	1.368	1.144
False	10296	119	.281	.253	1.504E-01	3.459E-01	.512	.634
False	16032	166	.204	.35	1.578E-01	1.092E-02	.921	1.052
False	5904	62	.076	.1	1.749E-01	1.267E-01	.86	.924
False	7664	85	.42	.435	1.798E-01	6.230E-01	.29	.41
False	16825	171	.045	.046	1.983E-01	8.754E-01	.659	.516
False	3434	18	.179	.188	2.089E-01	8.657E-01	1.071	1.702
False	16075	167	.343	.542	2.143E-01	2.502E-02	.992	1.001
False	7559	82	.53	.817	2.143E-01	2.085E-02	.84	.956
False	4747	36	.164	.158	2.162E-01	9.154E-01	1.829	1.88
False	4931	41	.139	.188	2.162E-01	6.297E-02	.778	.849
False	16079	168	.335	.525	2.198E-01	2.749E-02	1.001	1.003
False	7930	89	.606	.822	2.349E-01	8.814E-02	.893	.902
False	15113	156	2.263	3.425	2.387E-01	2.592E-02	.806	.963
False	4937	42	.163	.285	2.465E-01	4.899E-02	1.491	1.32
False	2512	2	.099	.085	2.811E-01	4.587E-01	1.066	1.189
False	20763	180	.027	.025	2.811E-01	3.576E-01	.379	.385
False	15865	164	2.481	3.431	2.938E-01	9.539E-02	.94	.996
False	11650	129	1.115	1.089	2.960E-01	8.717E-01	.718	.805
False	5530	55	.083	.084	3.047E-01	9.207E-01	.465	.403
False	5045	45	.419	.431	3.047E-01	8.196E-01	.514	.719
False	15120	157	2.735	3.768	3.134E-01	6.988E-02	.782	.939
False	15922	165	.392	.614	3.179E-01	4.469E-02	.884	1.177
False	24020	186	.018	.02	3.336E-01	3.843E-01	.567	.634
False	15127	158	2.84	3.724	3.358E-01	1.199E-01	.779	.938
False	4219	28	.055	.047	3.358E-01	4.704E-01	1.14	1.257
False	2553	4	.068	.068	3.404E-01	9.940E-01	.9	.818
False	4722	35	.064	.061	3.404E-01	8.248E-01	1.21	.942
False	15171	159	.363	.549	3.473E-01	4.884E-02	.866	1.112
False	7935	90	.587	.749	3.496E-01	1.740E-01	.915	.908
False	2783	7	.092	.052	3.542E-01	3.260E-02	1.386	.729
False	4964	43	1	2.64	3.612E-01	1.943E-02	2.135	1.742
False	7565	83	.506	.711	3.635E-01	7.343E-02	.836	1.004
False	8033	93	.139	.151	3.682E-01	5.353E-01	.679	.733
False	17879	172	.113	.09	3.729E-01	2.308E-01	1.082	.744
False	4784	38	.084	.087	3.753E-01	7.425E-01	.593	.625
False	3440	19	.168	.158	3.753E-01	8.814E-01	2.312	1.682
False	15821	163	.336	.571	3.848E-01	3.403E-02	1.045	1.268
False	14491	153	.103	.168	3.895E-01	1.164E-01	1.558	1.516
False	7242	79	.072	.084	3.991E-01	3.449E-01	.673	.874
False	4767	37	.149	.124	4.015E-01	3.770E-01	1.105	.931
False	22765	185	.024	.024	4.015E-01	8.755E-01	.548	.558
False	4672	34	.062	.108	4.088E-01	3.798E-02	.701	1.432

False	2272	0	.06	.05	4.136E-01	3.894E-01	1.196	.857
False	17923	174	.078	.065	4.209E-01	2.584E-01	.943	.641
False	5039	44	.357	.323	4.209E-01	4.854E-01	.714	.798
False	8404	96	.109	.103	4.258E-01	6.674E-01	.671	.55
False	3369	15	.205	.197	4.332E-01	9.284E-01	2.684	1.826
False	6181	65	.054	.045	4.356E-01	6.365E-01	2.166	1.614
False	11982	132	.116	.151	4.356E-01	1.894E-01	.854	1.11
False	2929	8	.073	.062	4.356E-01	3.570E-01	.974	.674
False	2935	9	.061	.107	4.356E-01	5.869E-02	.991	1.564
False	16779	169	.062	.061	4.430E-01	8.907E-01	.646	.666
False	18827	176	.016	.017	4.454E-01	5.578E-01	.419	.577
False	20915	181	.062	.055	4.578E-01	3.565E-01	.764	.552
False	9337	105	.076	.078	4.602E-01	8.857E-01	.746	.885
False	18405	175	.033	.032	4.677E-01	6.717E-01	.517	.392
False	2532	3	.082	.048	4.801E-01	7.495E-02	1.606	.953
False	11071	122	.427	.406	4.801E-01	7.523E-01	.924	.753
False	4281	30	.074	.075	4.801E-01	9.629E-01	.761	.838
False	16793	170	.064	.064	4.850E-01	9.583E-01	.57	.596
False	4917	40	.11	.117	4.925E-01	5.814E-01	.473	.637
False	17904	173	.097	.077	4.925E-01	1.910E-01	.983	.711
False	5171	50	.084	.258	4.975E-01	1.276E-02	2.246	1.811

Comparison of Far Normal (FN) vs. Near Margin Normal (NN) - Medulla Border (TM) in Tumor Grade IV

Data sets:
Near Margin Normal (NN) -- selected spectra from samples D528, D1344, D1347
Far Normal (FN) -- selected spectra from samples D528, D1344, D1347

Plot of Group Averages: <group 3TM FN in blue, group 3TM NN in red>



Plot of Group Differences:

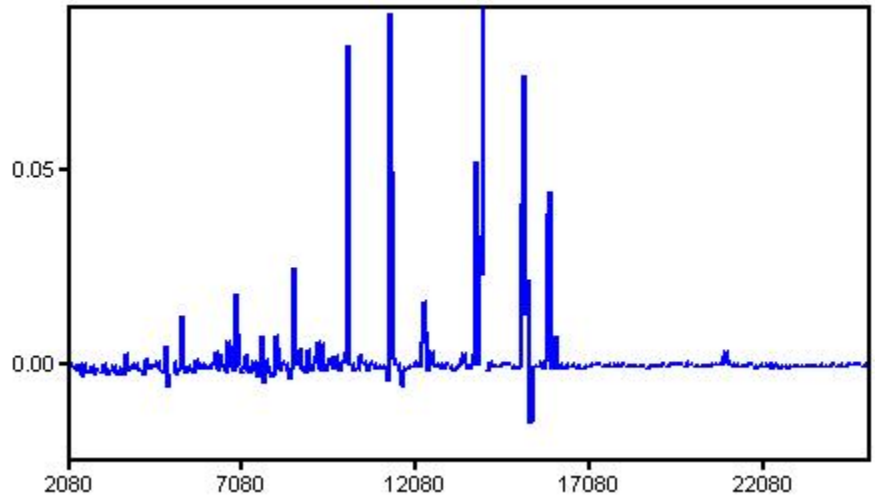


Table of features used in classifier and validation
(Selected features with Wil P < 0.00002)

MZ	ID	3TM FN Avg	3TM NN Avg	Wil P	T p	CV1	CV2
4893	28	.147	.044	6.991E-06	1.779E-04	1.298	.694
5357	40	.631	.15	6.278E-13	1.471E-09	.831	.372
5646	41	.582	.33	1.640E-08	3.165E-08	.476	.404
5933	47	.194	.07	2.755E-07	3.207E-06	.927	.499
6664	59	.225	.082	1.239E-07	6.015E-08	.765	.568
6725	60	.133	.055	7.637E-09	1.431E-06	.767	.76
7643	70	.349	.21	2.223E-12	9.376E-13	.263	.4
7798	74	.142	.028	3.275E-10	6.308E-08	.959	1.552
8422	81	.017	.096	3.901E-06	7.936E-07	1.774	1.086
8560	82	.628	.43	3.677E-06	6.207E-04	.413	.731
8588	84	.26	.152	2.017E-06	5.707E-05	.53	.842
8646	85	.199	.052	6.599E-06	3.027E-06	1.081	.567
8655	86	.166	.055	1.466E-07	6.077E-08	.819	.516
8957	88	.124	.056	1.326E-07	1.762E-09	.574	.416
8964	89	.124	.062	1.526E-08	2.248E-08	.571	.413
9246	91	.161	.05	8.253E-13	1.892E-09	.741	.57
9373	93	.262	.064	2.942E-07	2.681E-07	.99	.485
9615	95	.196	.048	1.916E-07	2.189E-08	.898	.555
9623	96	.19	.059	7.248E-07	1.548E-08	.8	.505
9772	98	.137	.052	8.069E-06	5.195E-06	.921	.505
10097	102	3.295	2	8.487E-07	3.303E-07	.42	.516
12273	126	.564	.137	5.656E-11	1.011E-07	.943	.815
12282	127	.548	.144	6.415E-12	1.542E-09	.786	.735
12501	129	.137	.022	5.990E-07	1.559E-06	1.191	1.118
13427	133	.127	.068	2.610E-08	9.314E-09	.501	.392
13440	134	.131	.064	1.420E-08	1.131E-08	.572	.404
15385	151	.219	.447	5.387E-06	1.250E-05	.304	.795
20925	167	.154	.067	6.945E-08	6.675E-09	.616	.468

K table : PDiff: 0; 3TM FN- group size: 53, 3TM NN- group size: 53

k	Errors	Misses	G1 Errors	G1 Misses	G2 Errors	G2 Misses
3	13	0	6	0	7	0
7	15	0	9	0	6	0
5	15	0	8	0	7	0
11	16	0	11	0	5	0
13	16	0	11	0	5	0
9	16	0	11	0	5	0
15	18	0	13	0	5	0
1	18	0	8	0	10	0
19	19	0	14	0	5	0
17	19	0	14	0	5	0

Complete feature table

Used	MZ	ID	3TM FN Avg	3TM NN Avg	Wil P	T p	CV1	CV2
True	5357	40	.631	.15	6.278E-13	1.471E-09	.831	.372
True	9246	91	.161	.05	8.253E-13	1.892E-09	.741	.57
True	7643	70	.349	.21	2.223E-12	9.376E-13	.263	.4
True	12282	127	.548	.144	6.415E-12	1.542E-09	.786	.735
True	12273	126	.564	.137	5.656E-11	1.011E-07	.943	.815
True	7798	74	.142	.028	3.275E-10	6.308E-08	.959	1.552
True	6725	60	.133	.055	7.637E-09	1.431E-06	.767	.76
True	13440	134	.131	.064	1.420E-08	1.131E-08	.572	.404
True	8964	89	.124	.062	1.526E-08	2.248E-08	.571	.413
True	5646	41	.582	.33	1.640E-08	3.165E-08	.476	.404
True	13427	133	.127	.068	2.610E-08	9.314E-09	.501	.392
True	20925	167	.154	.067	6.945E-08	6.675E-09	.616	.468
True	6664	59	.225	.082	1.239E-07	6.015E-08	.765	.568
True	8957	88	.124	.056	1.326E-07	1.762E-09	.574	.416
True	8655	86	.166	.055	1.466E-07	6.077E-08	.819	.516
True	9615	95	.196	.048	1.916E-07	2.189E-08	.898	.555
True	5933	47	.194	.07	2.755E-07	3.207E-06	.927	.499
True	9373	93	.262	.064	2.942E-07	2.681E-07	.99	.485
True	12501	129	.137	.022	5.990E-07	1.559E-06	1.191	1.118
True	9623	96	.19	.059	7.248E-07	1.548E-08	.8	.505
True	10097	102	3.295	2	8.487E-07	3.303E-07	.42	.516
True	8588	84	.26	.152	2.017E-06	5.707E-05	.53	.842
True	8560	82	.628	.43	3.677E-06	6.207E-04	.413	.731
True	8422	81	.017	.096	3.901E-06	7.936E-07	1.774	1.086
True	15385	151	.219	.447	5.387E-06	1.250E-05	.304	.795
True	8646	85	.199	.052	6.599E-06	3.027E-06	1.081	.567
True	4893	28	.147	.044	6.991E-06	1.779E-04	1.298	.694
True	9772	98	.137	.052	8.069E-06	5.195E-06	.921	.505
False	6133	49	.099	.046	2.144E-05	6.056E-06	.775	.604
False	5350	39	.745	.134	2.810E-05	3.963E-06	1.218	.742
False	13385	132	.088	.063	2.810E-05	3.326E-05	.322	.49
False	2391	0	.014	.132	3.212E-05	6.258E-04	1.396	1.841
False	8016	78	.262	.076	4.413E-05	4.259E-06	1.044	.703
False	5054	33	.119	.351	1.053E-04	4.759E-05	.54	1.119
False	4931	29	.126	.236	1.163E-04	3.676E-05	.596	.72
False	3478	16	.042	.131	1.348E-04	6.424E-04	1.042	1.365
False	8570	83	.926	.708	1.524E-04	6.088E-02	.471	1.009
False	4887	27	.127	.041	1.720E-04	2.738E-06	.965	.834
False	9367	92	.266	.066	1.762E-04	7.383E-06	1.153	.59
False	13153	131	.022	.047	1.941E-04	1.689E-04	.927	.911
False	3368	11	.051	.231	2.036E-04	9.675E-04	1.945	1.61
False	15267	148	.925	.625	3.919E-04	1.676E-05	.239	.689
False	12348	128	.311	.2	4.010E-04	5.481E-04	.584	.674
False	7173	67	.08	.032	5.028E-04	3.546E-05	.929	.9
False	8713	87	.122	.038	5.142E-04	4.264E-05	1.114	1.096
False	10088	101	2.891	1.878	6.860E-04	6.164E-04	.585	.652
False	5042	32	.324	.247	7.012E-04	1.299E-03	.357	.506
False	9960	99	.162	.114	7.820E-04	2.416E-03	.582	.51
False	3439	14	.097	.324	7.991E-04	2.727E-03	1.477	1.603
False	3427	12	.22	.122	7.991E-04	5.413E-03	.958	1.118
False	13907	138	.704	.469	8.166E-04	2.546E-03	.548	.843
False	6657	58	.234	.083	8.901E-04	2.757E-05	1.054	.579

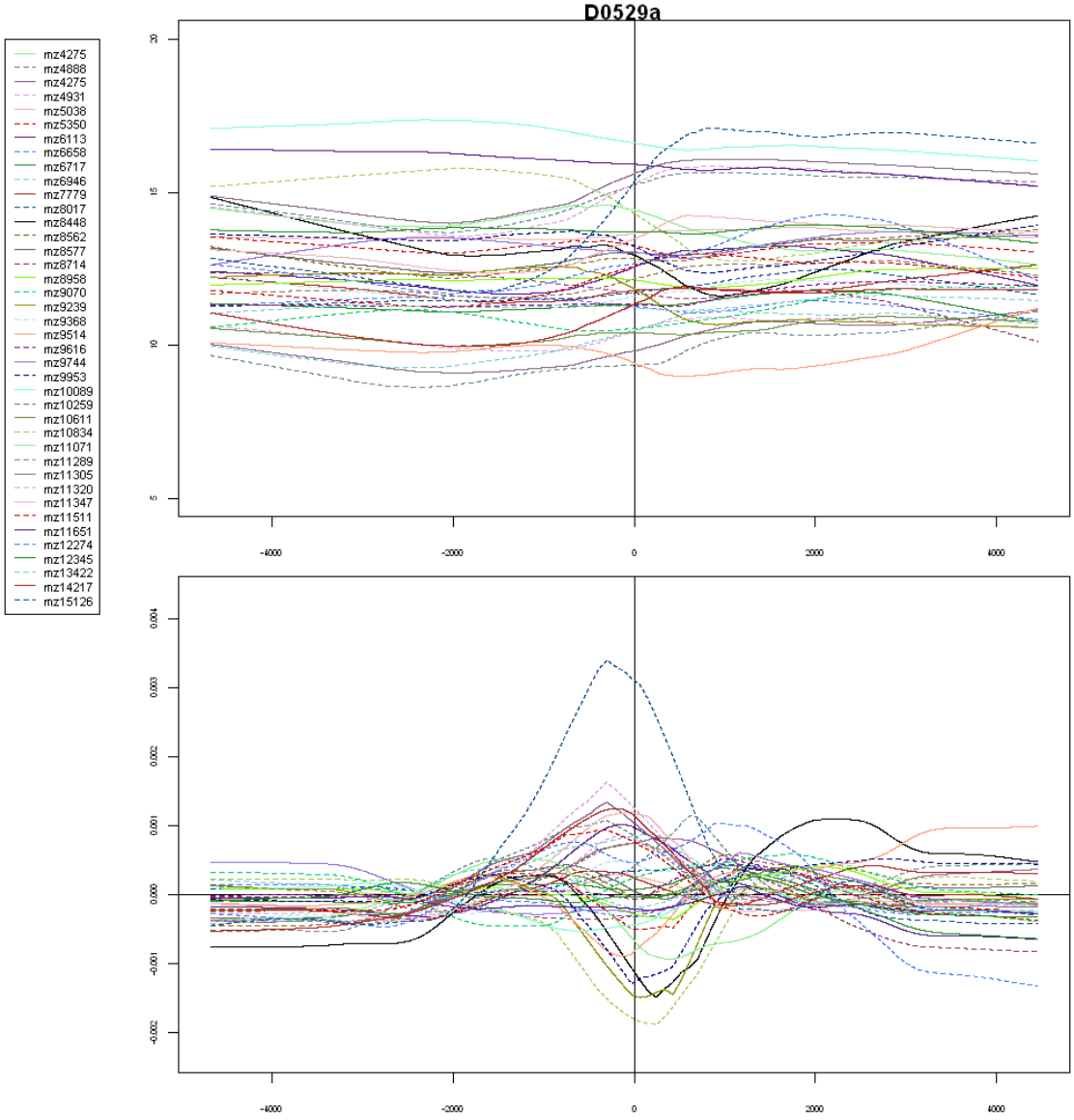
False	9751	97	.136	.078	9.492E-04	2.064E-03	.867	.845
False	10304	107	.29	.181	9.906E-04	1.651E-04	.62	.528
False	4766	26	.058	.084	1.148E-03	2.802E-03	.432	.648
False	8349	80	.069	.03	1.198E-03	3.767E-05	.9	.693
False	10133	103	.262	.192	1.999E-03	1.852E-03	.497	.49
False	5825	46	.087	.121	3.520E-03	5.308E-03	.611	.55
False	3434	13	.086	.188	3.725E-03	2.012E-03	.768	1.187
False	5654	42	.831	.597	4.168E-03	3.073E-03	.561	.527
False	5069	35	.953	.526	4.407E-03	2.340E-01	2.662	1.015
False	11303	115	2.887	2.445	5.013E-03	2.002E-02	.275	.453
False	3362	10	.056	.14	5.199E-03	3.794E-04	.869	1.129
False	14631	145	.063	.067	5.295E-03	5.942E-01	.774	.405
False	11312	116	3.581	3.034	6.451E-03	4.474E-02	.342	.504
False	16793	158	.044	.056	6.924E-03	2.706E-02	.593	.45
False	9522	94	.096	.036	6.924E-03	3.182E-04	1.174	.974
False	8039	79	.14	.096	7.172E-03	3.898E-03	.652	.623
False	10267	104	.252	.164	7.299E-03	7.578E-04	.664	.486
False	10294	106	.298	.194	8.829E-03	1.696E-03	.7	.553
False	6179	50	.047	.114	8.829E-03	2.021E-03	.921	1.302
False	5062	34	.304	.574	9.292E-03	2.544E-02	1.68	1.22
False	10846	109	.161	.126	9.943E-03	2.747E-02	.51	.631
False	19746	165	.029	.037	1.063E-02	2.223E-01	1.185	.799
False	10077	100	1.11	.756	1.081E-02	1.107E-01	1.212	1.153
False	16506	157	.051	.062	1.428E-02	1.292E-01	.747	.565
False	12657	130	.056	.068	1.474E-02	8.529E-02	.576	.544
False	22550	171	.06	.067	1.596E-02	7.135E-01	1.914	.949
False	4937	30	.205	.297	1.596E-02	6.495E-02	1.071	.954
False	17542	160	.014	.022	1.622E-02	1.337E-02	.916	.79
False	4747	25	.063	.143	1.647E-02	1.408E-02	1.393	1.515
False	20176	166	.033	.039	1.867E-02	9.077E-02	.528	.469
False	6324	53	.158	.106	1.955E-02	3.813E-04	.604	.356
False	15292	149	1.095	.838	2.415E-02	8.331E-04	.171	.612
False	5085	36	.15	.135	2.451E-02	7.274E-01	1.878	.994
False	6953	64	.2	.159	2.717E-02	9.726E-02	.591	.817
False	4566	24	.504	.392	2.757E-02	1.614E-01	.859	.982
False	3325	9	.171	.088	3.463E-02	2.202E-02	1.382	1.2
False	12231	125	.172	.21	3.561E-02	4.899E-01	.821	1.786
False	7682	73	.169	.239	3.611E-02	5.958E-03	.454	.691
False	7930	75	.563	.424	3.712E-02	7.029E-02	.778	.796
False	15339	150	.978	1.159	3.712E-02	6.833E-02	.33	.549
False	4012	19	.056	.085	3.764E-02	4.281E-02	.471	1.186
False	2476	1	.046	.119	3.764E-02	6.071E-04	.92	1.209
False	9192	90	.13	.144	4.085E-02	6.943E-01	.851	1.549
False	15853	152	1.768	1.786	4.987E-02	9.591E-01	.911	1.141
False	11345	117	2.146	2.005	5.321E-02	4.384E-01	.355	.535
False	21543	168	.019	.023	5.459E-02	5.830E-02	.556	.594
False	22213	170	.038	.035	5.745E-02	2.995E-01	.401	.414
False	7662	71	.332	.307	6.042E-02	2.757E-01	.299	.43
False	2929	5	.05	.073	6.429E-02	2.314E-02	.592	.889
False	5170	37	.056	.188	6.429E-02	9.204E-04	1.43	1.432
False	5668	43	.374	.336	7.090E-02	2.010E-01	.395	.467
False	17204	159	.04	.033	8.081E-02	7.259E-02	.558	.514
False	15130	146	3.713	3.261	8.369E-02	4.723E-01	.859	.999
False	15871	153	2.582	2.296	8.967E-02	5.251E-01	.892	1.009
False	10284	105	.249	.169	9.384E-02	9.934E-03	.803	.546
False	3901	18	.103	.086	9.597E-02	2.564E-01	.75	.918
False	14016	140	2.367	2.072	9.705E-02	3.484E-01	.633	.828

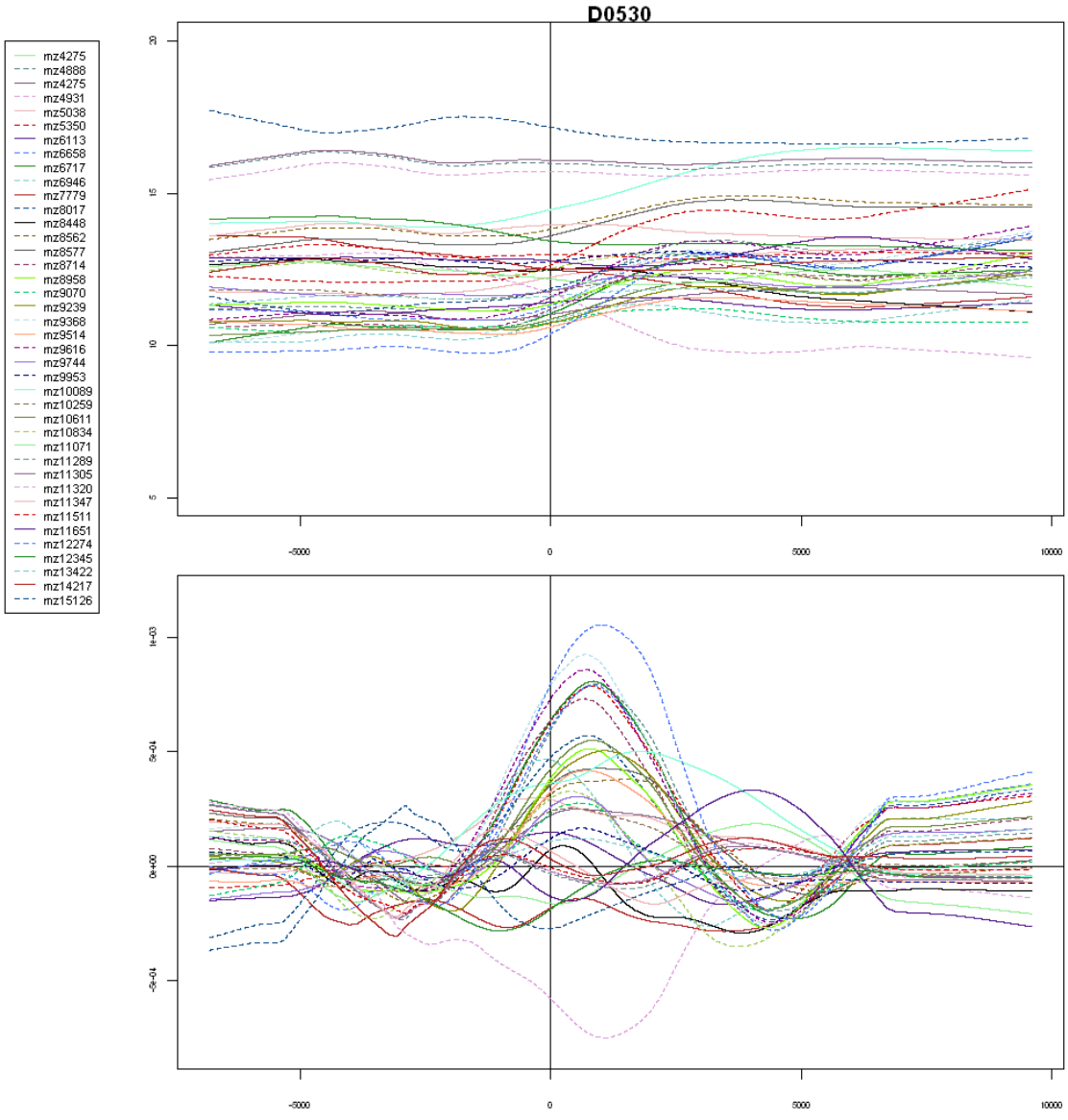
False	6626	57	.058	.153	9.924E-02	1.172E-03	1.223	1.277
False	15882	154	2.215	1.946	9.924E-02	4.712E-01	.873	.979
False	10832	108	.173	.142	9.924E-02	2.408E-01	.813	.958
False	7937	76	.595	.411	1.003E-01	3.719E-02	.948	.703
False	4559	23	.452	.446	1.003E-01	9.537E-01	.966	1.249
False	4964	31	.549	1.979	1.014E-01	3.733E-04	1.424	1.375
False	4282	22	.144	.157	1.083E-01	5.661E-01	.443	.96
False	14230	143	.182	.183	1.119E-01	9.899E-01	.557	.829
False	6881	61	.25	.297	1.131E-01	7.654E-02	.434	.53
False	13789	137	2.547	2.238	1.155E-01	3.660E-01	.649	.826
False	6435	55	.172	.138	1.323E-01	1.512E-01	.773	.748
False	7559	68	.612	.476	1.323E-01	1.895E-01	.981	.936
False	11385	119	.736	.663	1.478E-01	3.000E-01	.509	.516
False	15173	147	.582	.649	1.537E-01	5.630E-01	.755	1.102
False	7953	77	.138	.142	1.613E-01	8.622E-01	1.005	.717
False	14222	142	.19	.198	1.692E-01	7.390E-01	.491	.752
False	15928	155	.423	.479	1.822E-01	5.096E-01	.839	1.035
False	6115	48	.134	.113	1.822E-01	1.659E-01	.62	.625
False	4046	20	.072	.077	1.856E-01	6.657E-01	.931	.742
False	11078	112	.241	.233	1.856E-01	7.713E-01	.505	.723
False	7566	69	.714	.522	1.907E-01	1.025E-01	1.029	.82
False	18579	164	.02	.023	1.924E-01	2.434E-01	.588	.694
False	13982	139	.421	.405	2.048E-01	6.891E-01	.406	.598
False	14489	144	.047	.052	2.102E-01	5.101E-01	.818	.86
False	11654	122	.589	.528	2.120E-01	3.703E-01	.62	.621
False	6428	54	.128	.172	2.157E-01	1.393E-01	.965	1.006
False	6226	51	.075	.088	2.232E-01	4.609E-01	1.142	1.045
False	3471	15	.081	.236	2.269E-01	1.674E-01	1.012	3.418
False	11354	118	2.117	2.155	2.405E-01	8.699E-01	.458	.656
False	16079	156	.348	.355	2.444E-01	9.155E-01	.922	1.07
False	11069	111	.222	.229	2.504E-01	8.083E-01	.582	.782
False	23695	172	.023	.026	2.525E-01	2.968E-01	.504	.587
False	11395	120	.599	.559	2.773E-01	5.533E-01	.582	.615
False	2935	6	.115	.123	2.880E-01	7.494E-01	1.069	1.08
False	10894	110	.044	.054	2.923E-01	2.879E-01	.887	1.046
False	3722	17	.161	.104	2.945E-01	1.273E-01	1.48	1.184
False	5696	45	.141	.154	2.966E-01	3.379E-01	.404	.504
False	2780	4	.024	.033	2.988E-01	2.809E-01	1.134	1.593
False	11609	121	.12	.116	3.077E-01	7.366E-01	.48	.545
False	3273	8	.121	.12	3.166E-01	9.601E-01	.785	.718
False	5263	38	.116	.118	3.166E-01	8.312E-01	.455	.652
False	13760	136	1.185	1.354	3.189E-01	2.179E-01	.428	.629
False	6911	63	.214	.192	3.211E-01	5.470E-01	.907	.926
False	17929	163	.063	.056	3.441E-01	3.718E-01	.733	.491
False	6895	62	.564	.543	3.487E-01	7.795E-01	.655	.761
False	7008	65	.623	.587	3.534E-01	6.724E-01	.649	.801
False	2957	7	.061	.082	3.605E-01	2.078E-01	.94	1.27
False	11822	124	.179	.166	3.796E-01	5.973E-01	.741	.657
False	6312	52	.086	.101	3.917E-01	4.215E-01	.926	1.138
False	7029	66	.135	.138	4.014E-01	8.633E-01	.686	.752
False	13724	135	.235	.251	4.186E-01	6.022E-01	.589	.703
False	7669	72	.314	.326	4.211E-01	6.782E-01	.41	.513
False	14060	141	.389	.373	4.211E-01	7.778E-01	.696	.818
False	4115	21	.085	.079	4.260E-01	4.742E-01	.541	.543
False	6549	56	.12	.124	4.359E-01	8.547E-01	.737	.725
False	11814	123	.164	.154	4.459E-01	6.840E-01	.879	.796
False	11269	114	.52	.486	4.459E-01	5.644E-01	.642	.563

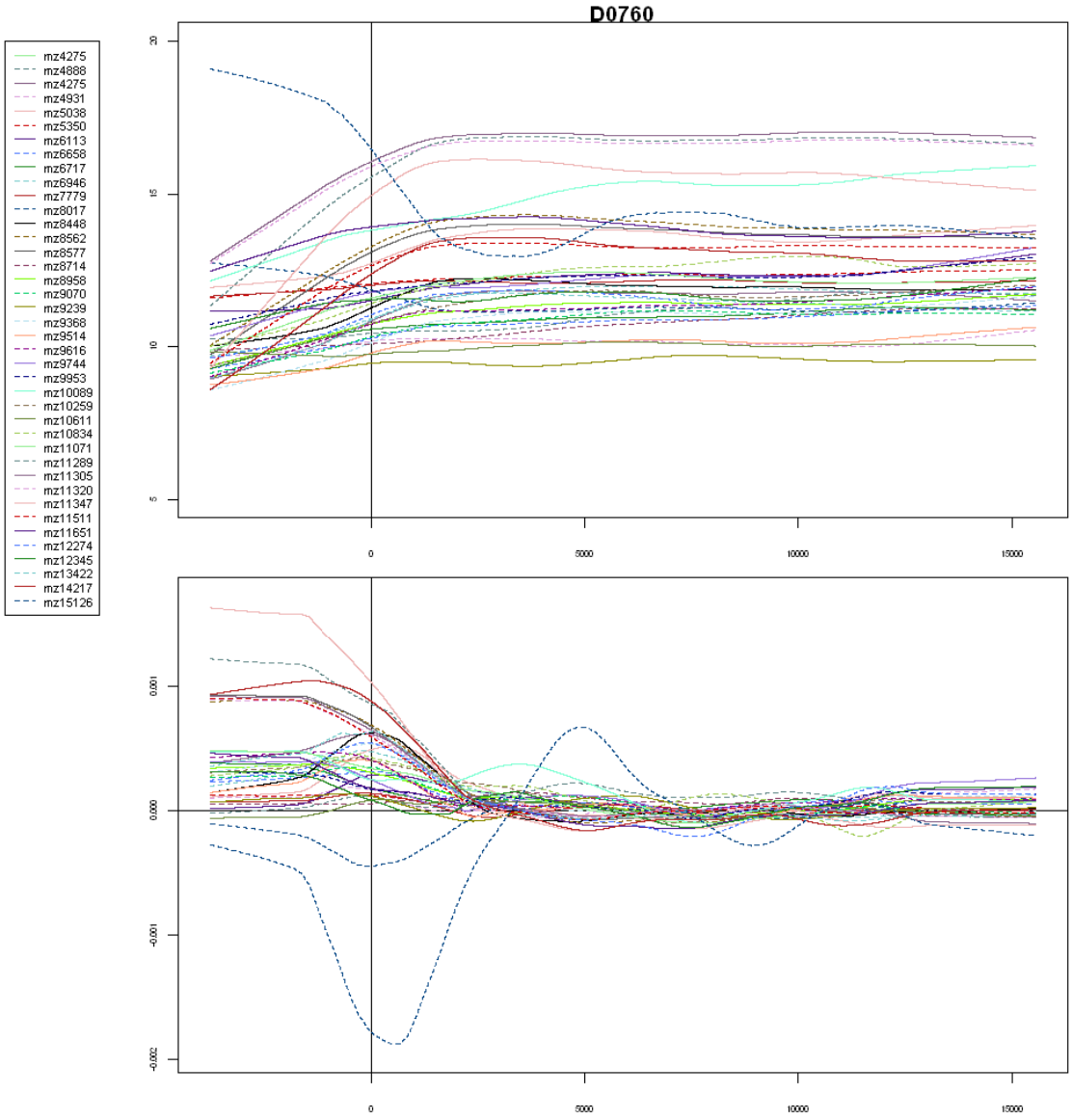
False	24758	173	.022	.025	4.534E-01	3.986E-01	.538	.774
False	2750	3	.453	.542	4.660E-01	6.258E-01	1.62	2.03
False	2744	2	.085	.092	4.685E-01	6.368E-01	.868	.983
False	11259	113	.369	.344	4.685E-01	5.771E-01	.701	.586
False	17918	162	.07	.061	4.811E-01	3.107E-01	.777	.547
False	17881	161	.082	.069	4.811E-01	2.080E-01	.832	.557
False	5675	44	.381	.406	4.811E-01	5.908E-01	.514	.652
False	21778	169	.042	.053	4.886E-01	1.245E-01	.533	.921

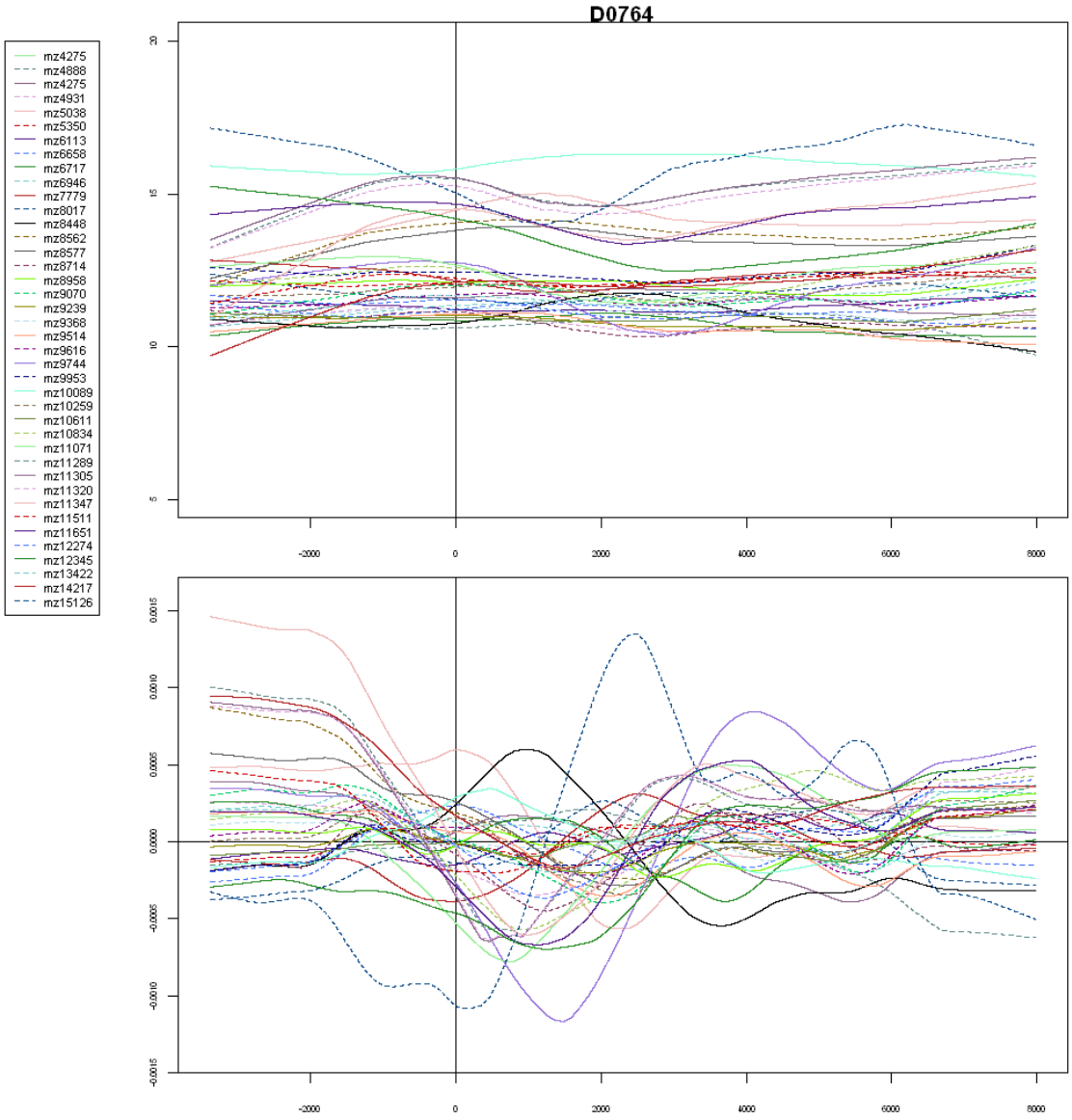
Appendix C

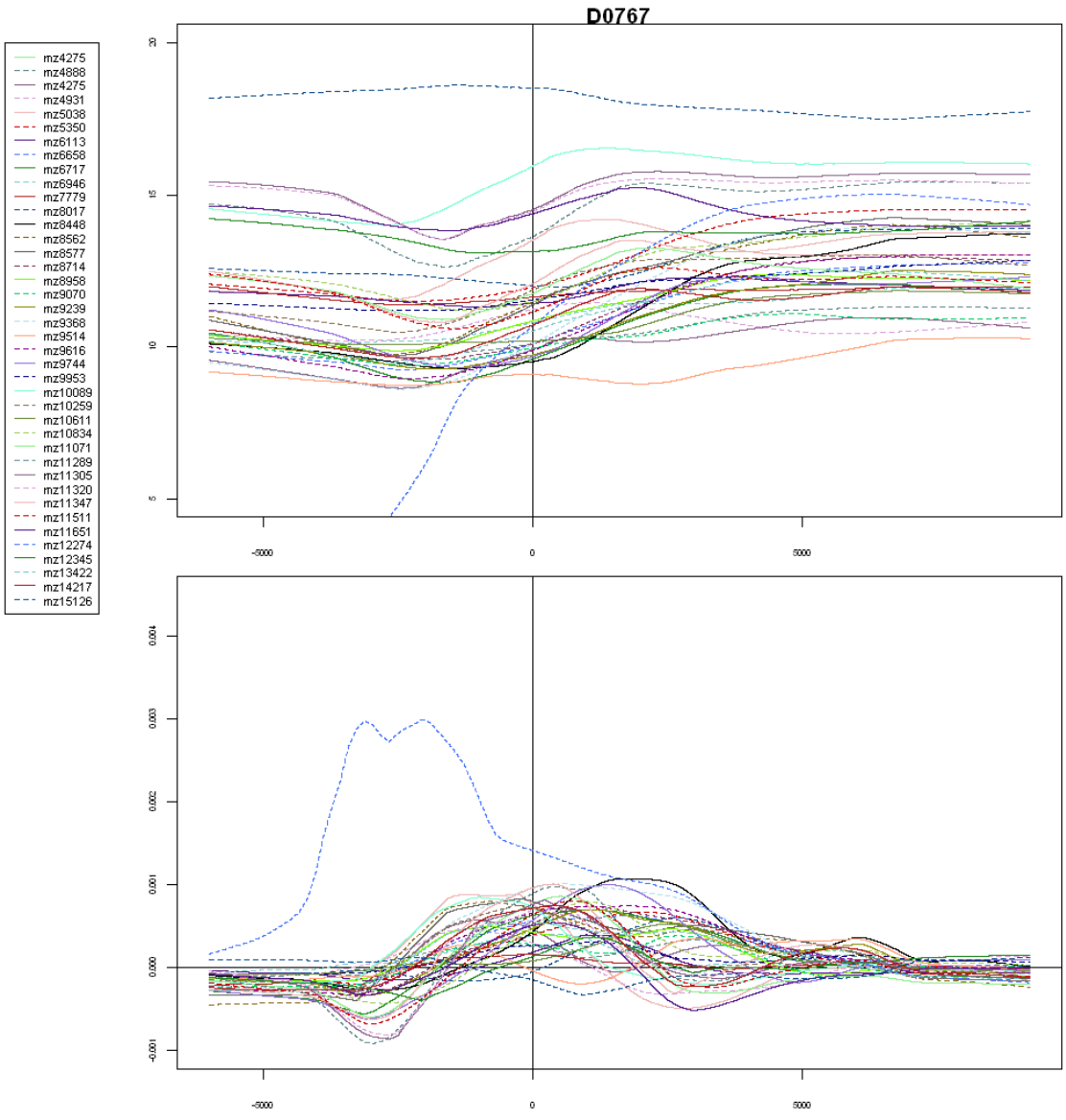
Note: Refer to Figure 32 for graph and axes information pertaining to all figures in Appendix C.

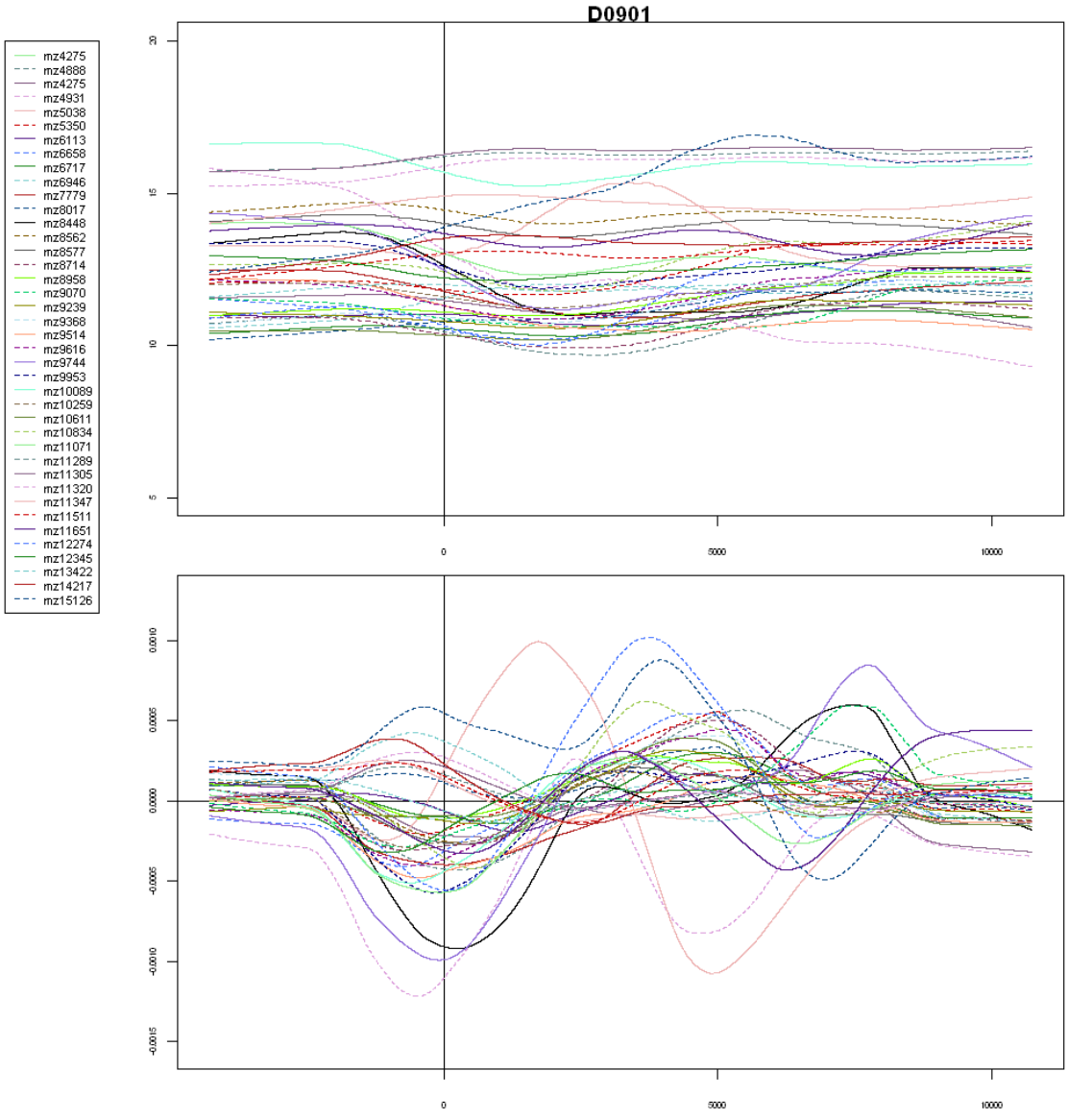


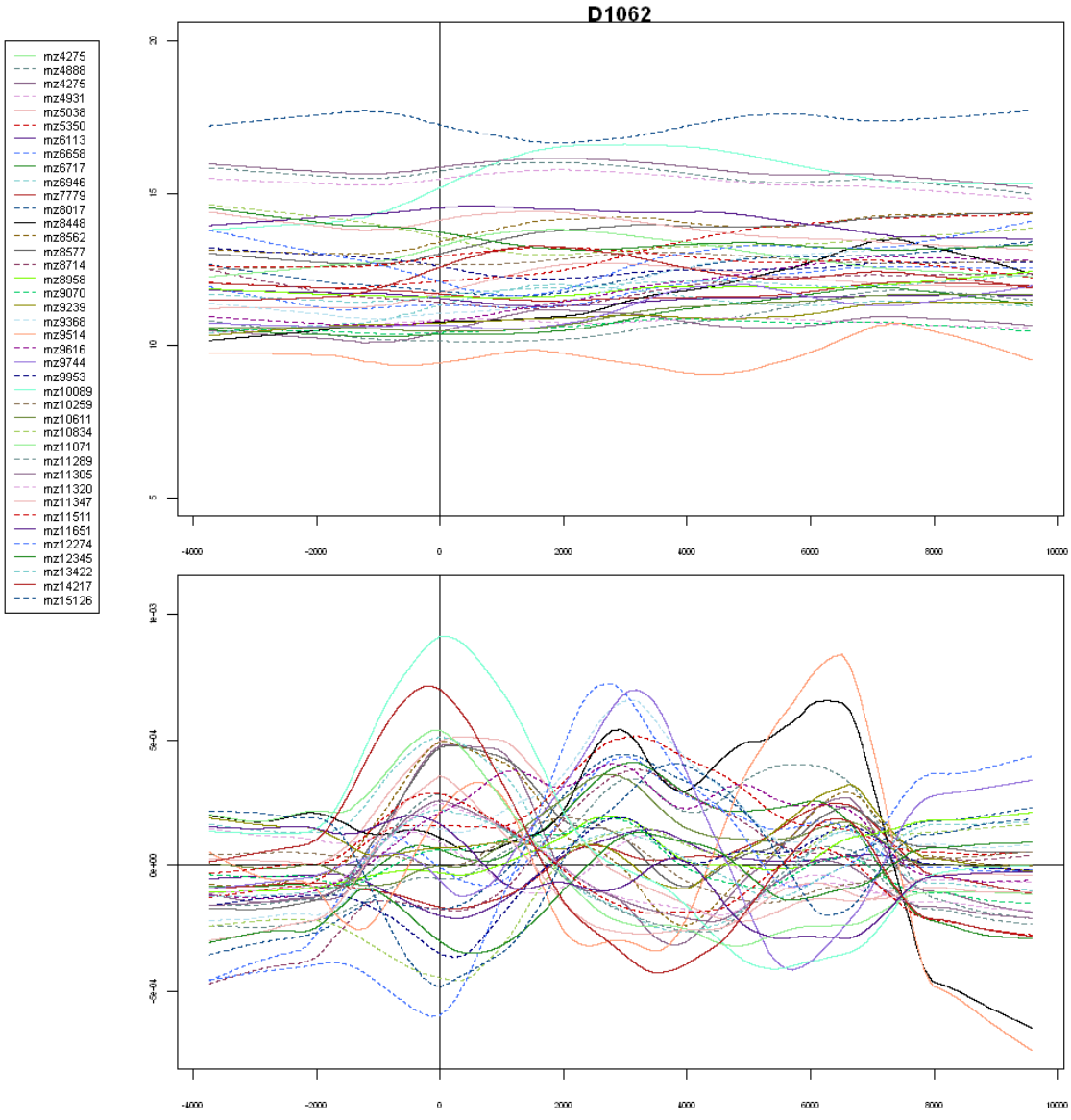

D0530
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-5000
0
5000
10000
1e-03
5e-04
0e+00
-5e-04
-5000
0
5000
10000

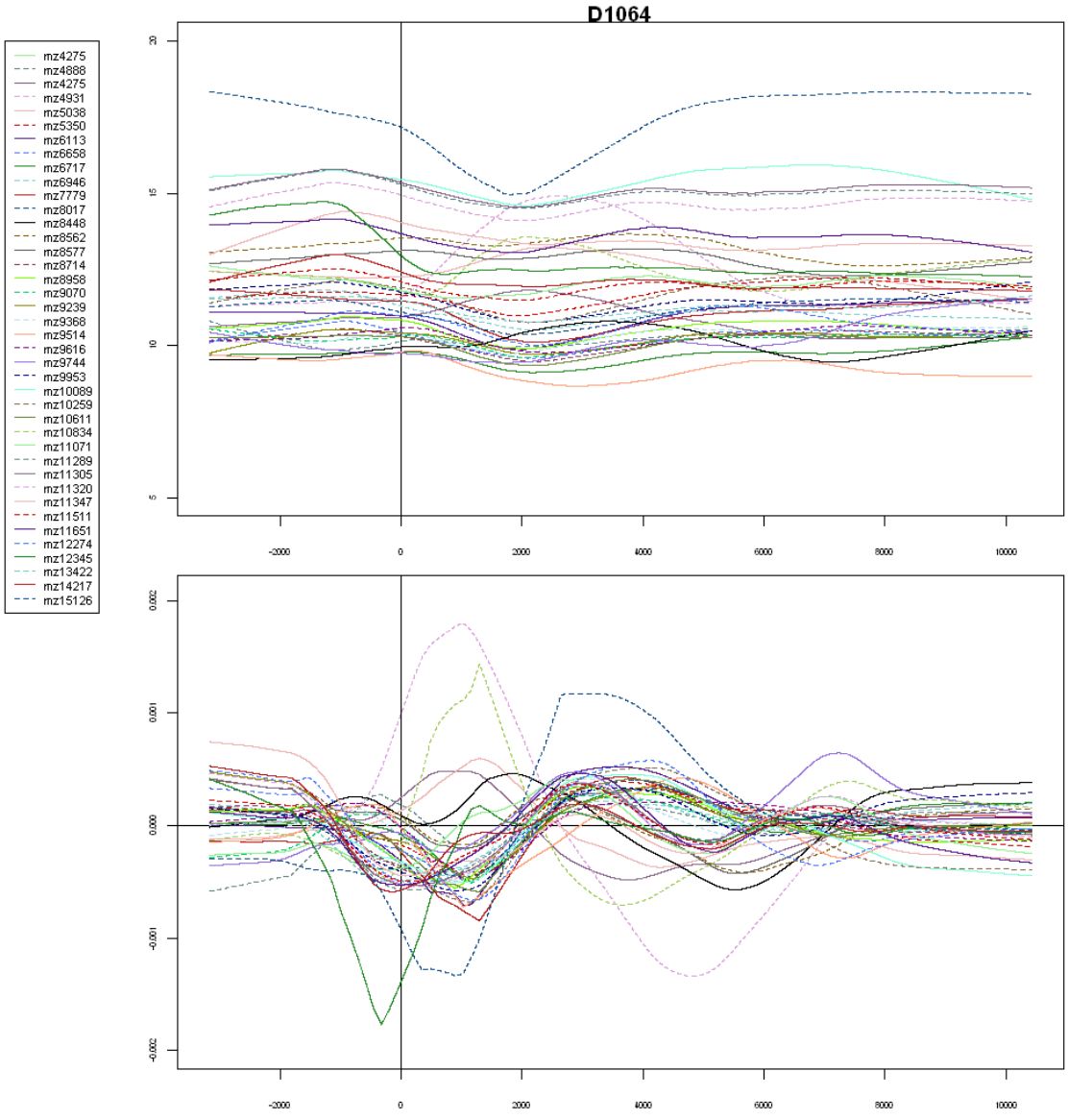

D0760
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
0
5000
10000
15000
0.001
0.000
-0.001
-0.002
0
5000
10000
15000

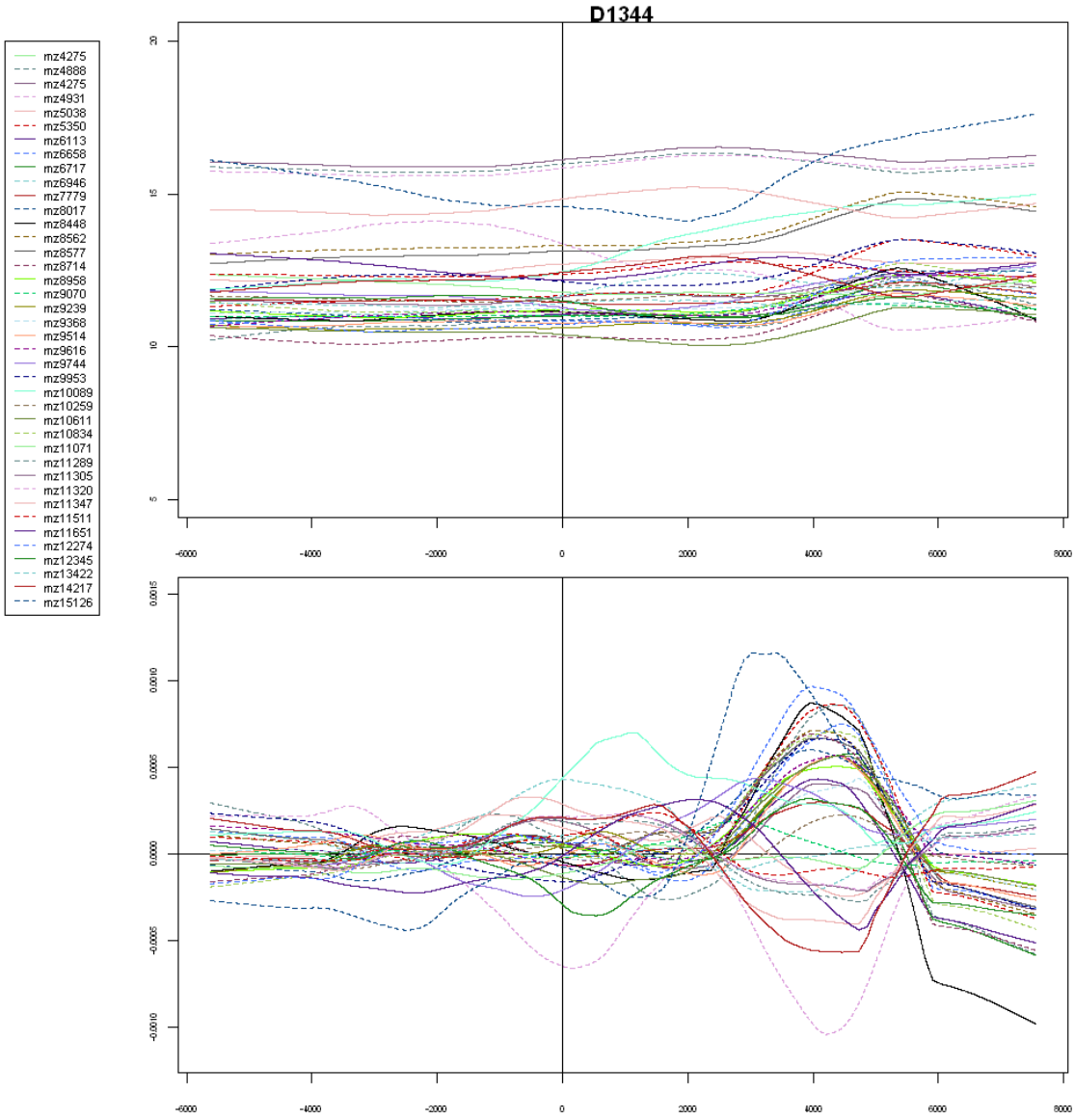

D0764
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-2000
0
2000
4000
6000
8000
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015
-2000
0
2000
4000
6000
8000

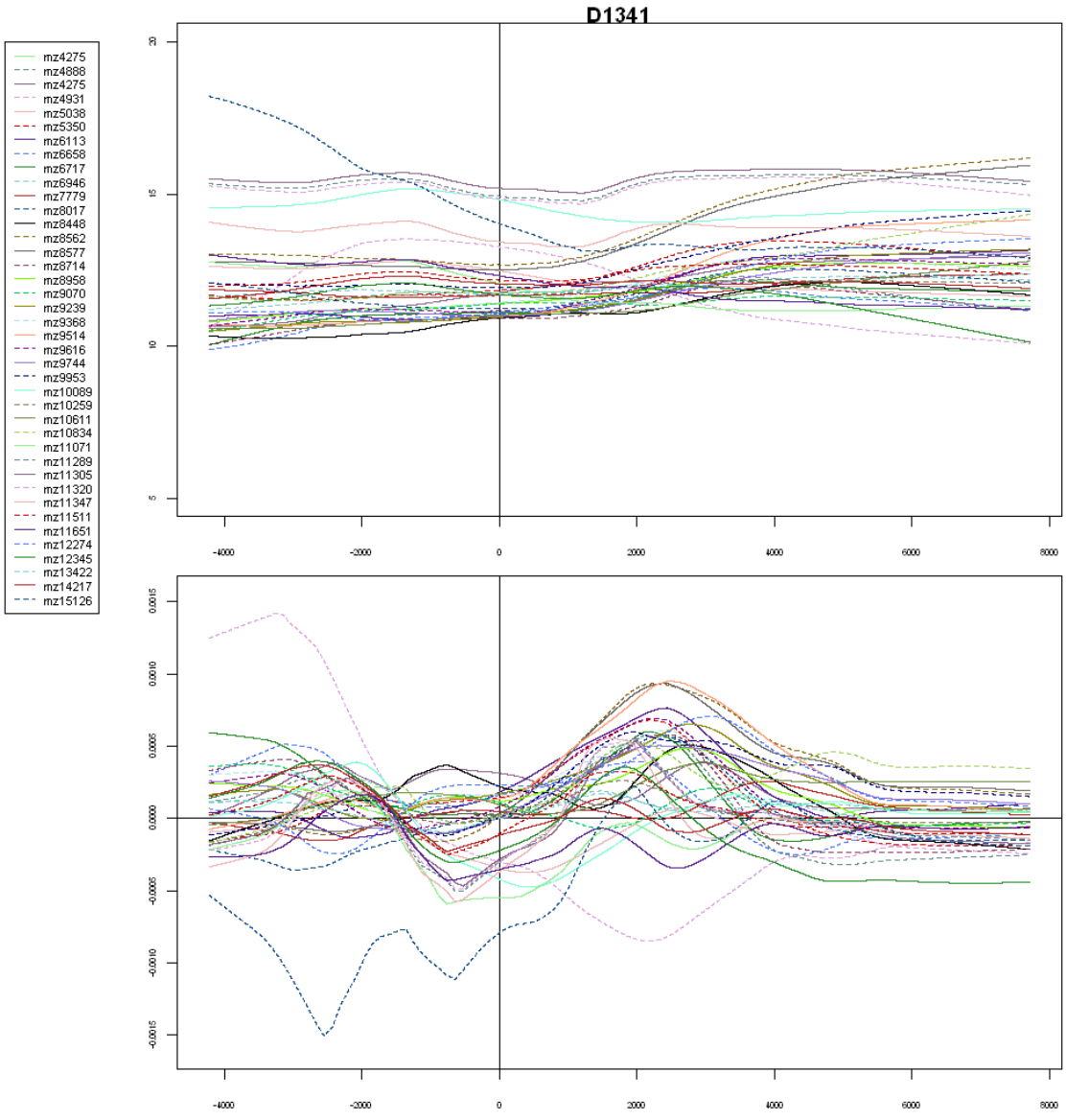

D0767
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-5000
0
5000
0.004
0.003
0.002
0.001
0.000
-0.001
-5000
0
5000

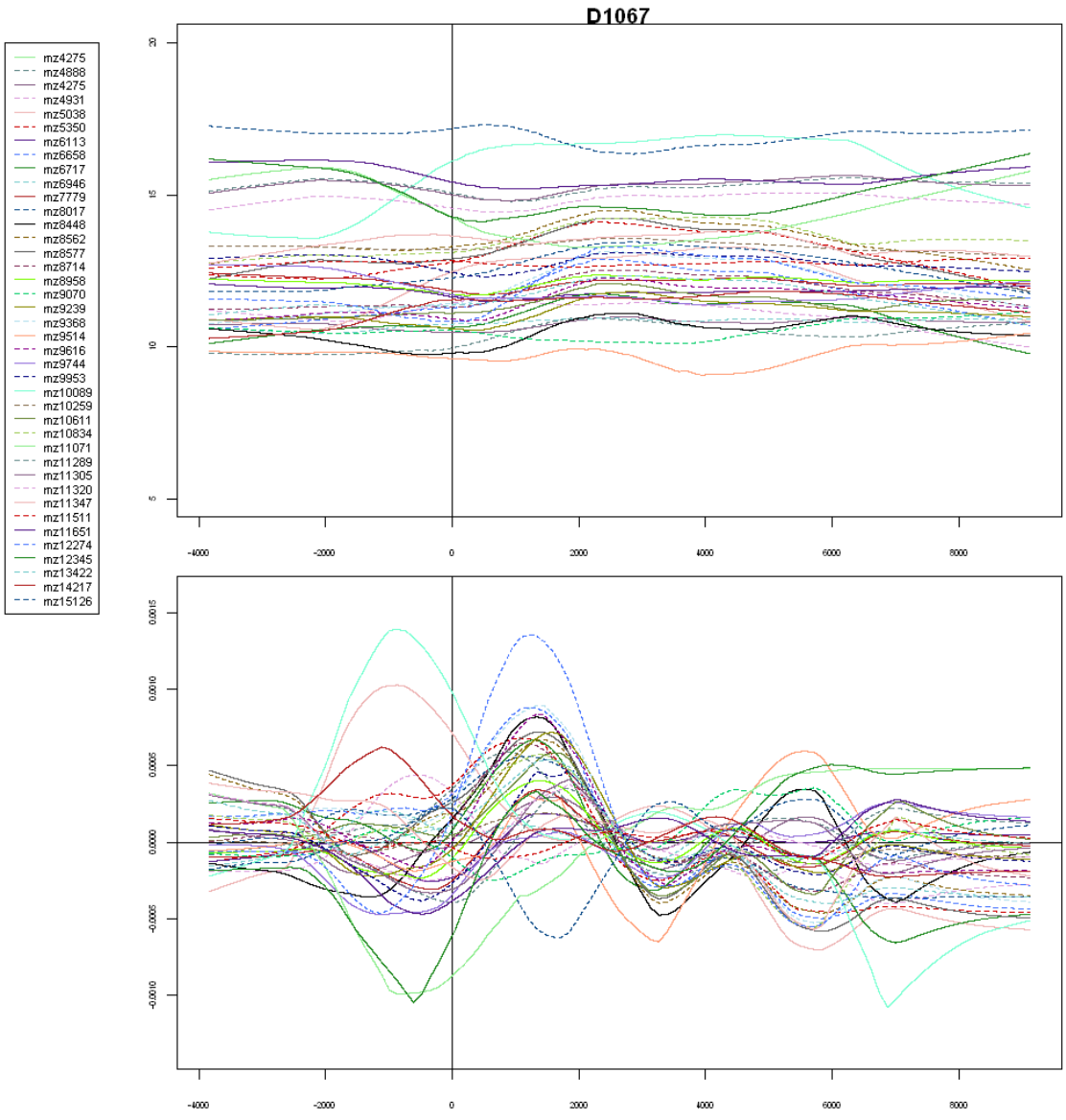

D0901
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
0
5000
10000
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015
0
5000
10000

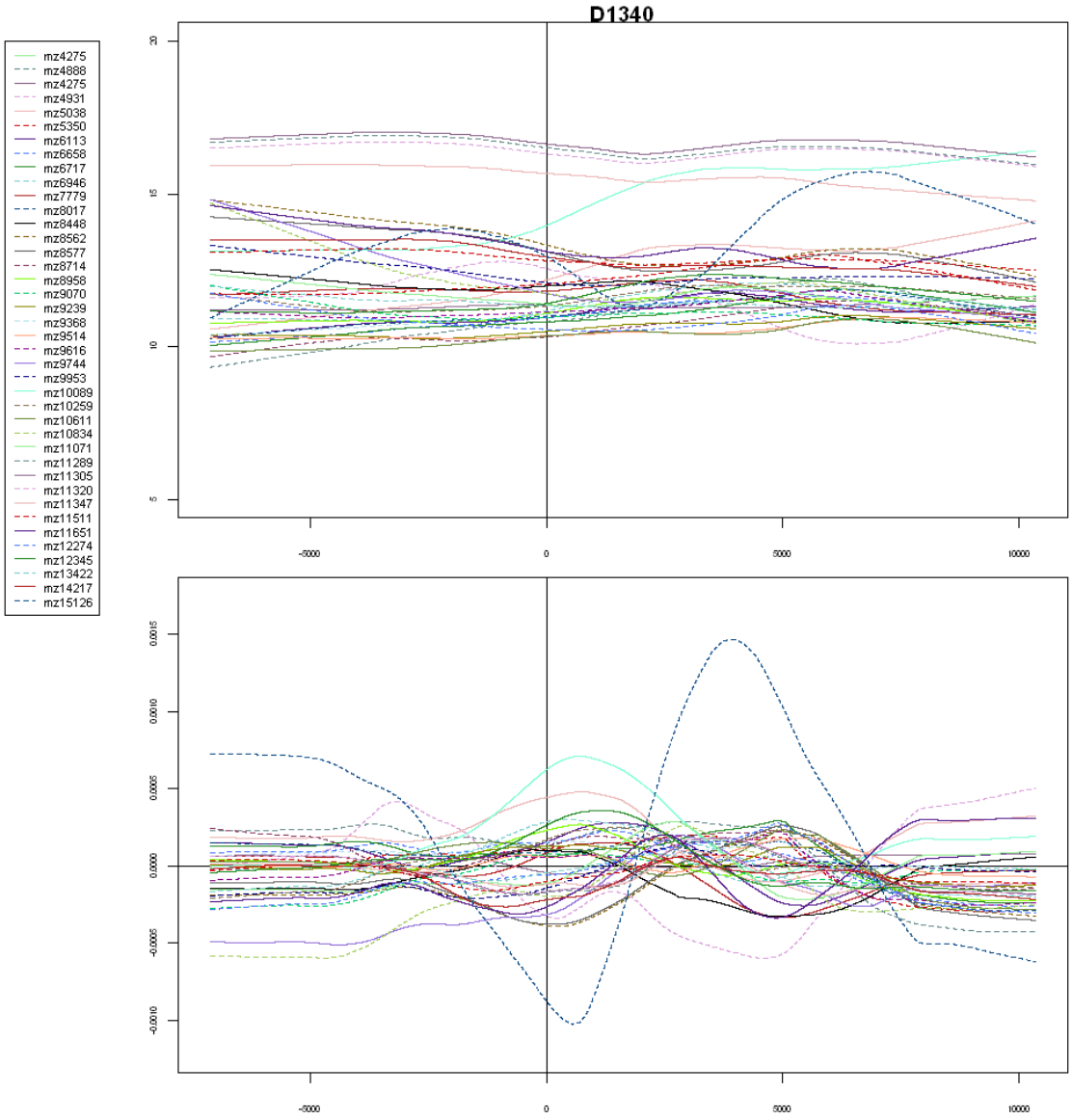

D1062
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
5
10
15
20
-4000
-2000
0
2000
4000
6000
8000
10000
-5e-04
0e+00
5e-04
1e-03

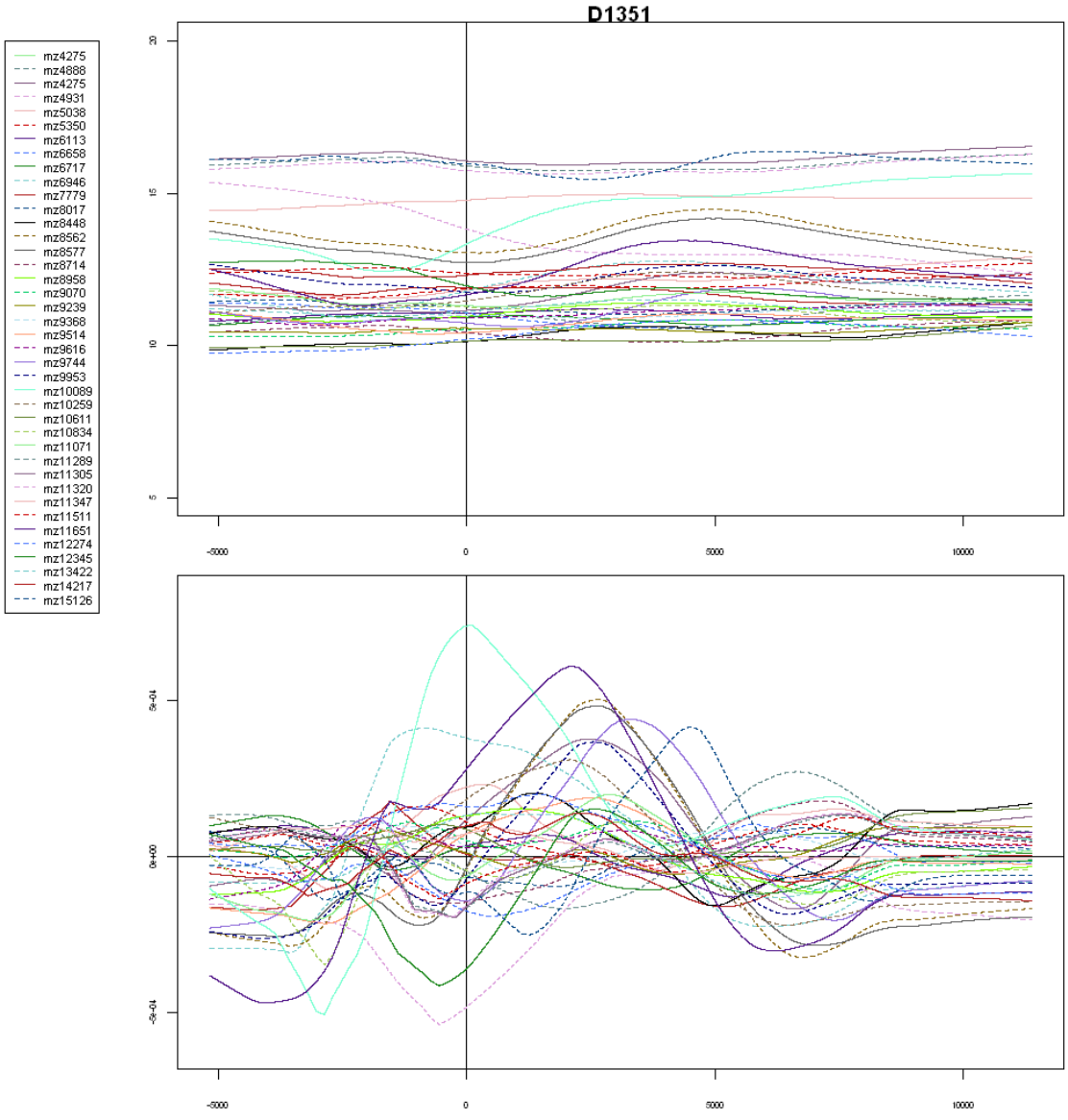

D1064
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-2000
0
2000
4000
6000
8000
10000
0.002
0.001
0.000
-0.001
-0.002
-2000
0
2000
4000
6000
8000
10000

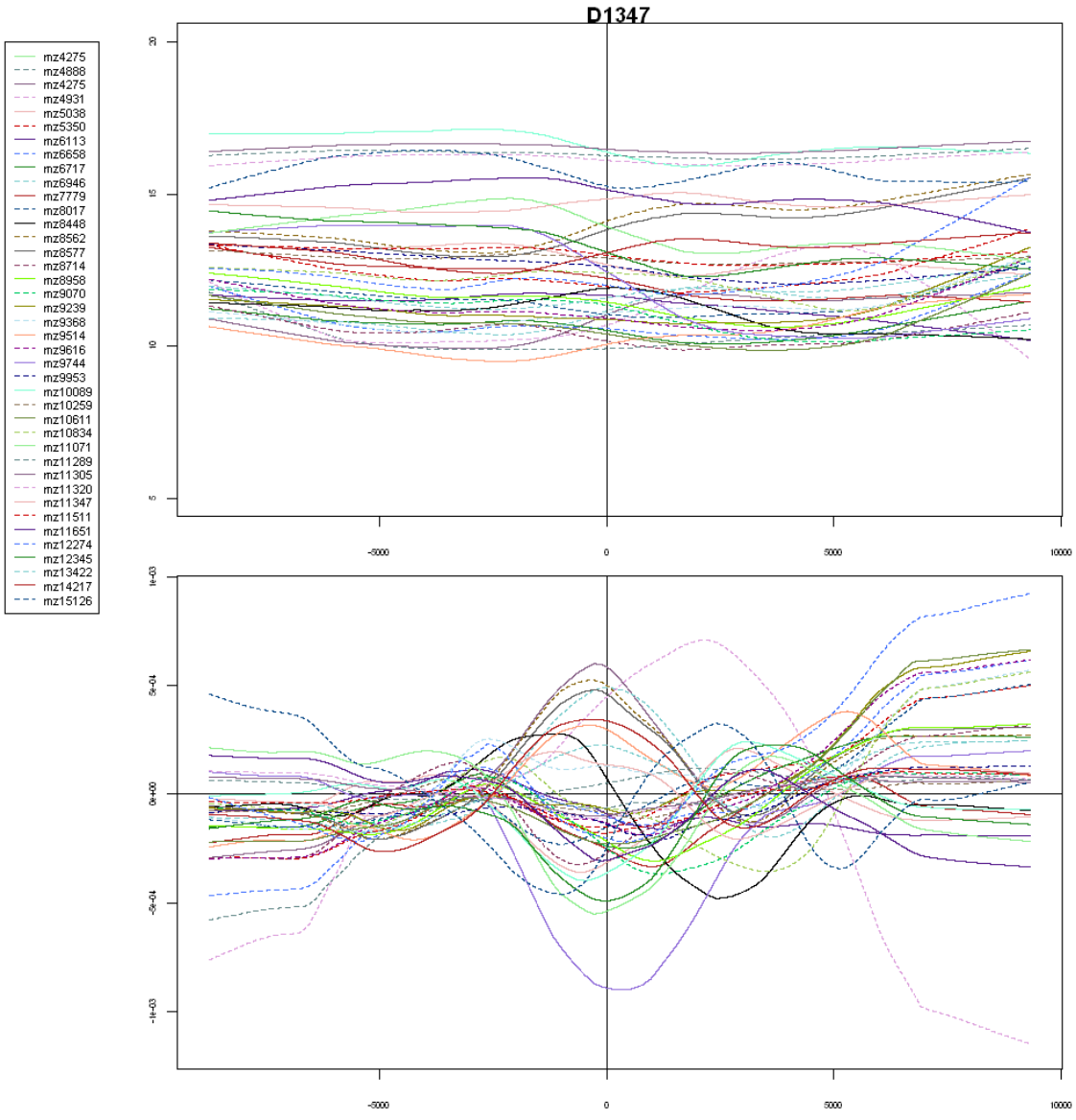

D1344
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126

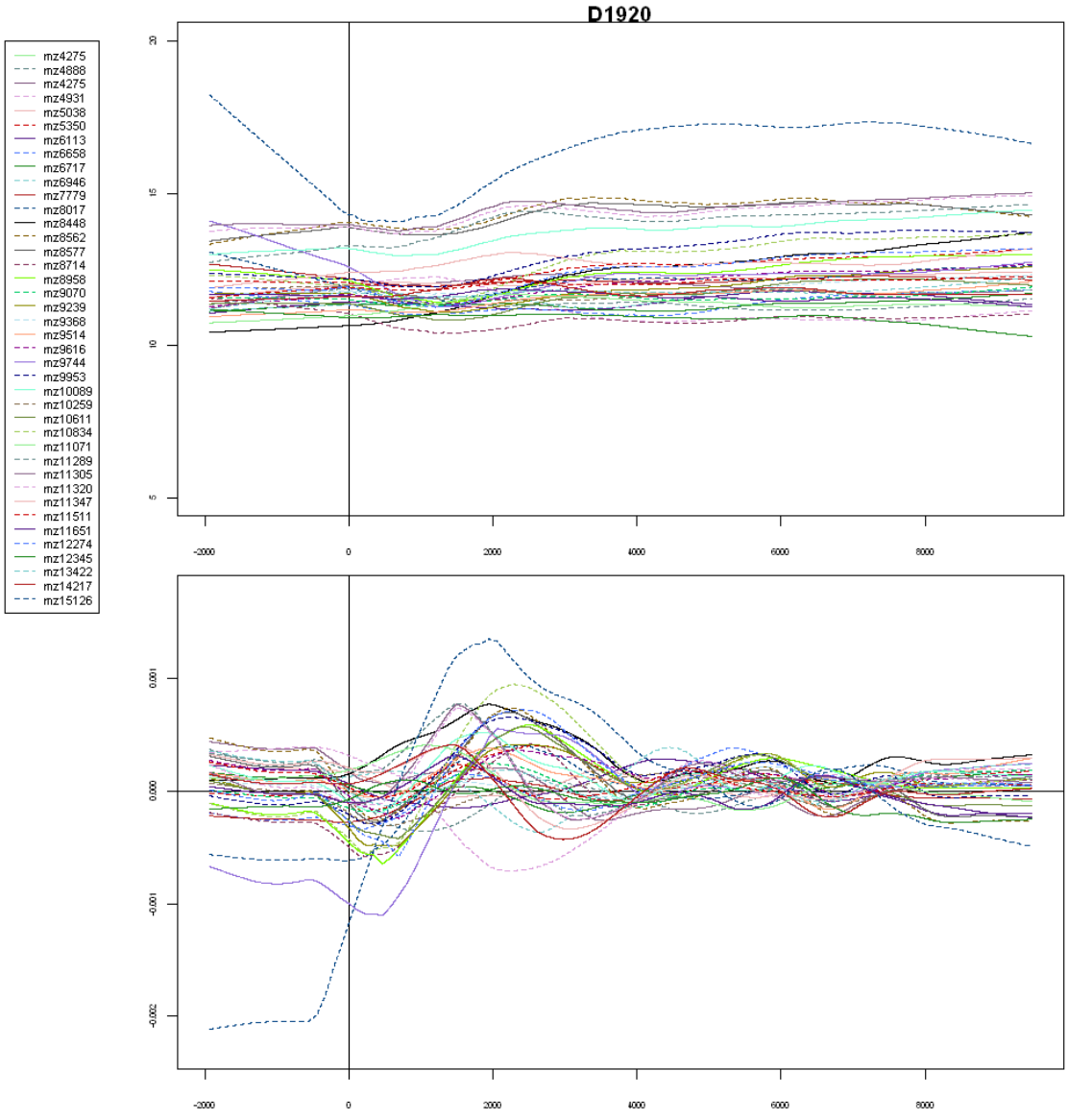

D1341
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-4000
-2000
0
2000
4000
6000
8000
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015

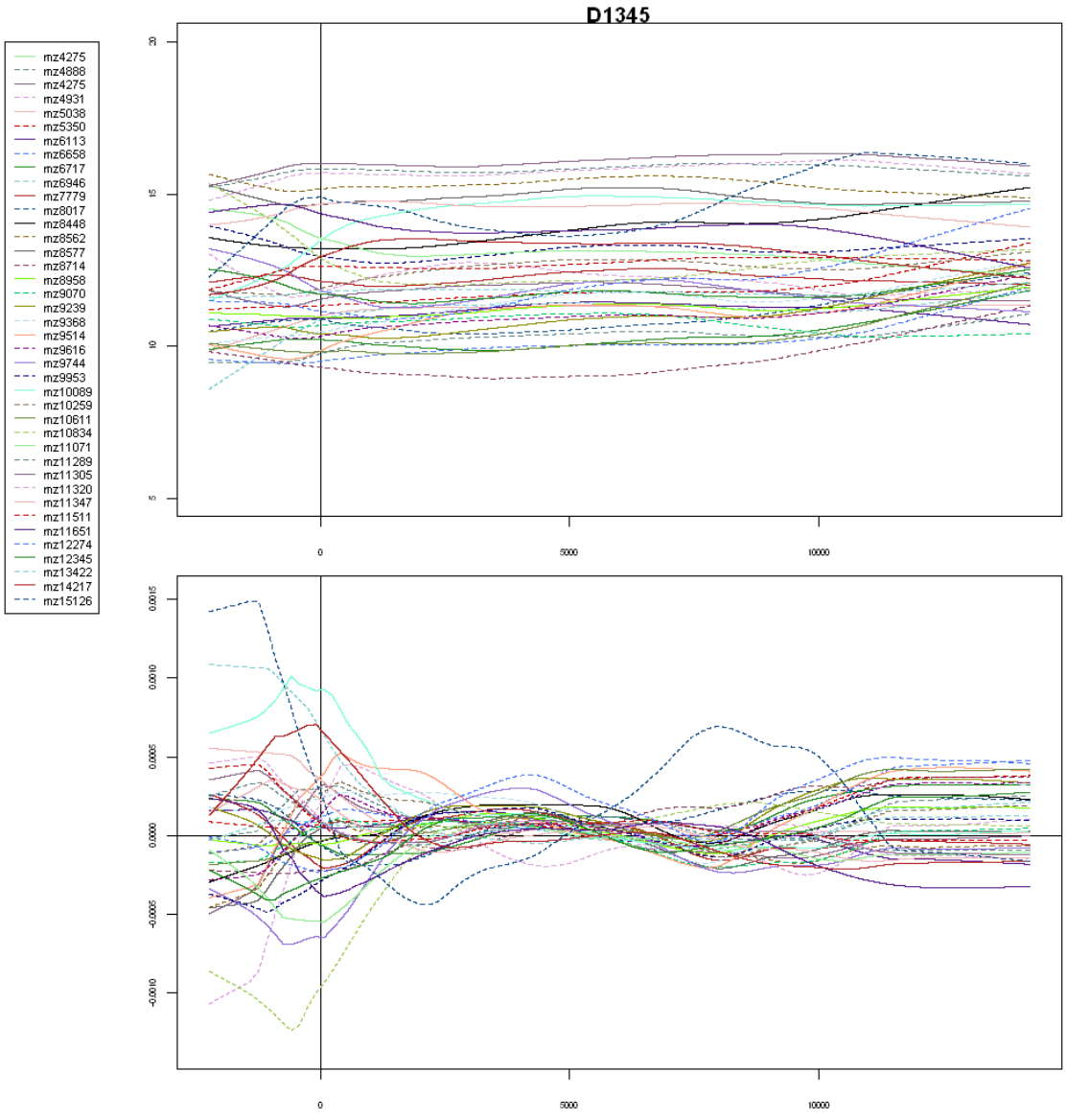

D1067
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-4000
-2000
0
2000
4000
6000
8000
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010

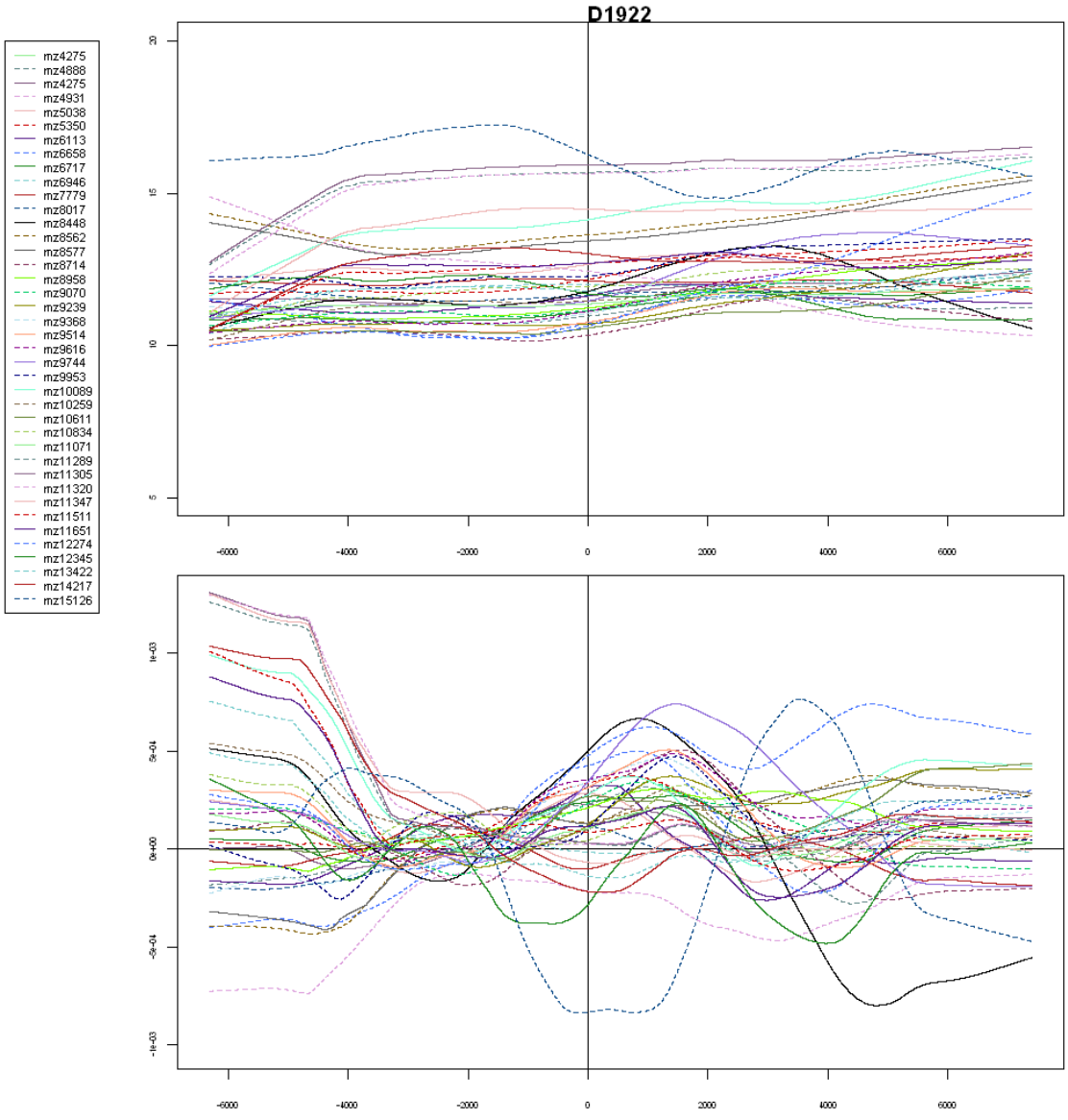

D1340
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-5000
0
5000
10000
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-5000
0
5000
10000

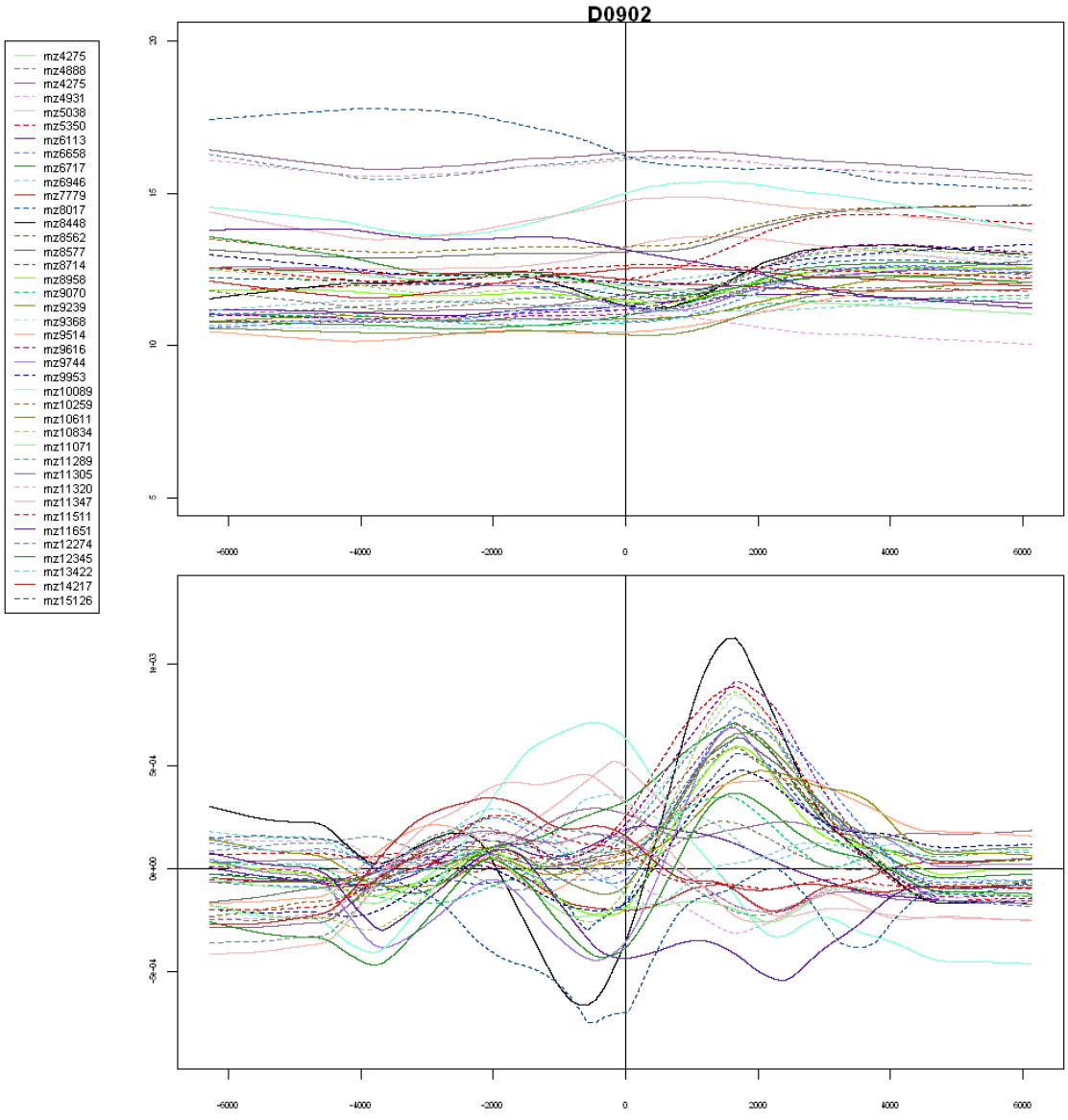

D1351
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-5000
0
5000
10000
5e+04
0e+00
-5e+04
-5000
0
5000
10000

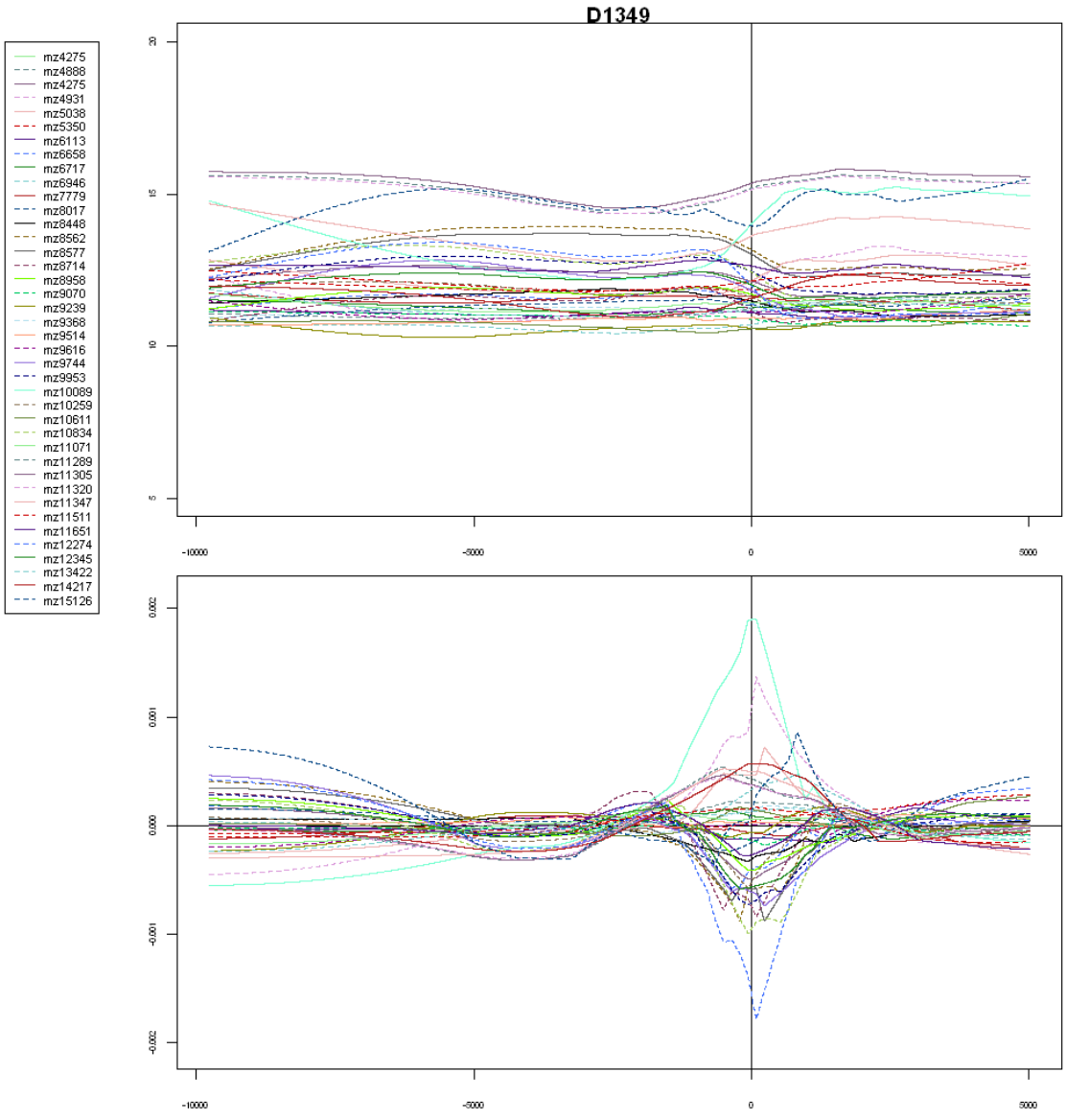

D1347
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-5000
0
5000
10000
1e-03
5e-04
0e+00
-5e-04
-1e-03
-5000
0
5000
10000

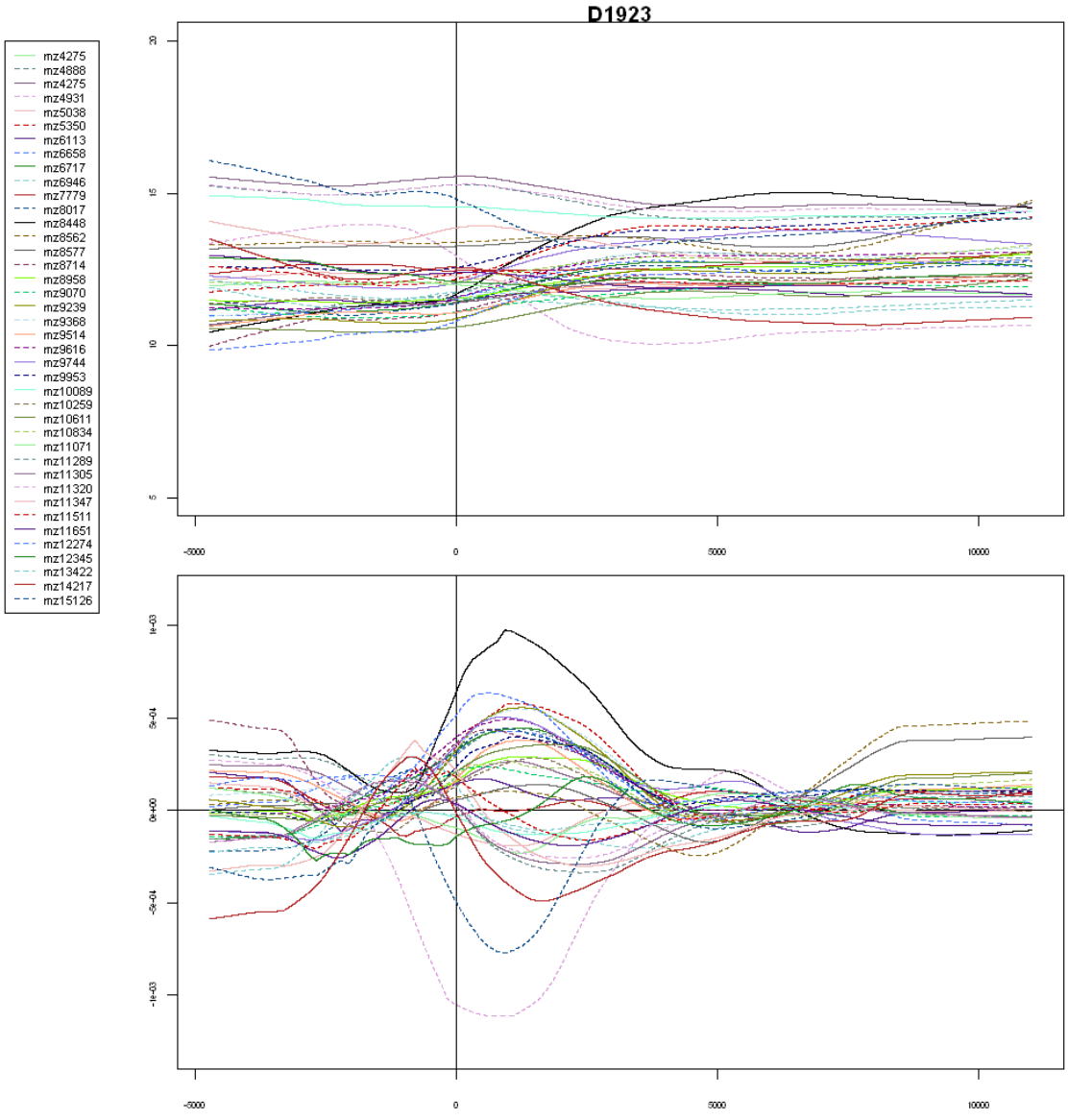

D1920
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-2000
0
2000
4000
6000
8000
0.001
0.000
-0.001
-0.002
-2000
0
2000
4000
6000
8000

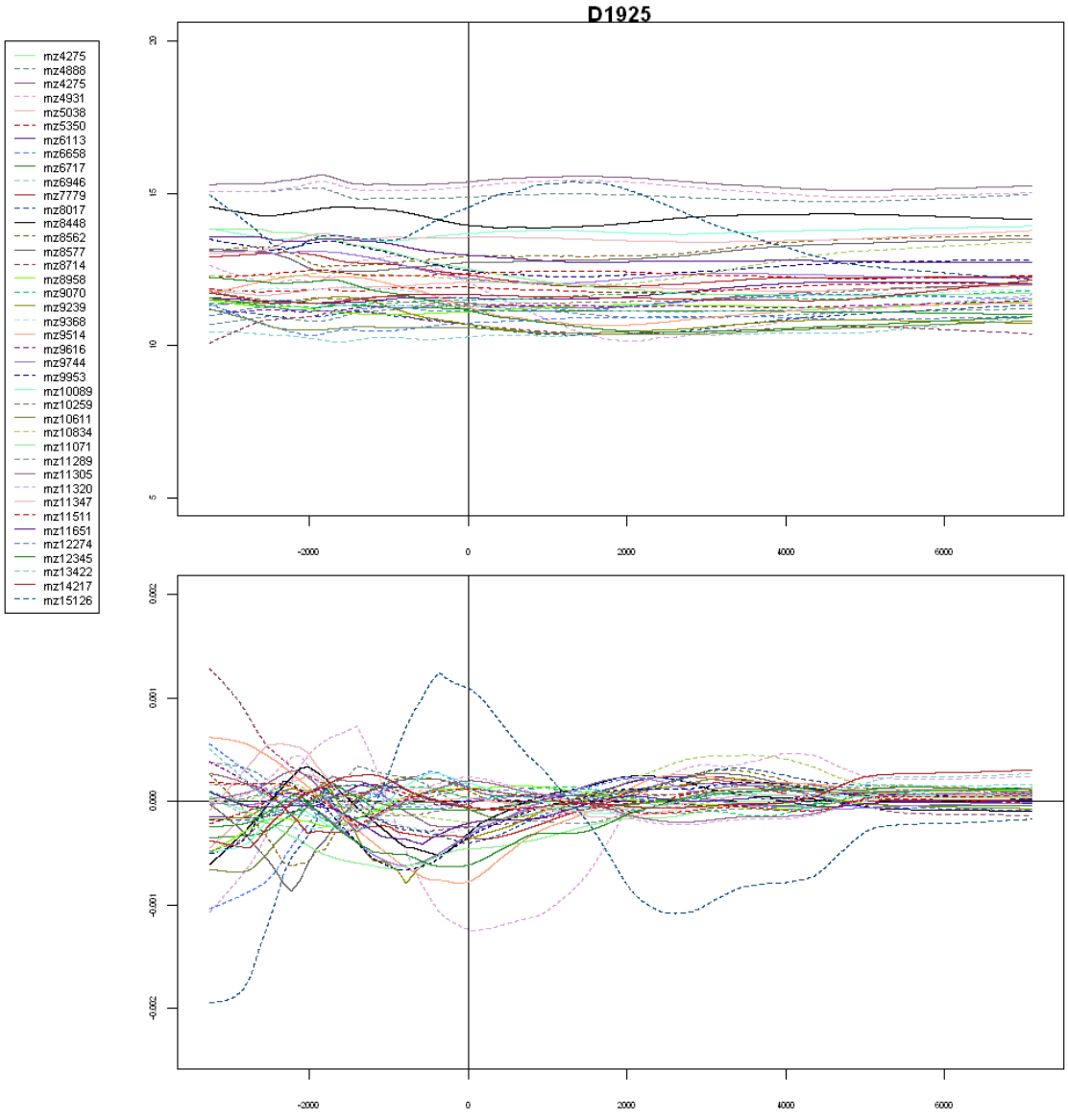

D1345
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
0
5000
10000
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
0
5000
10000

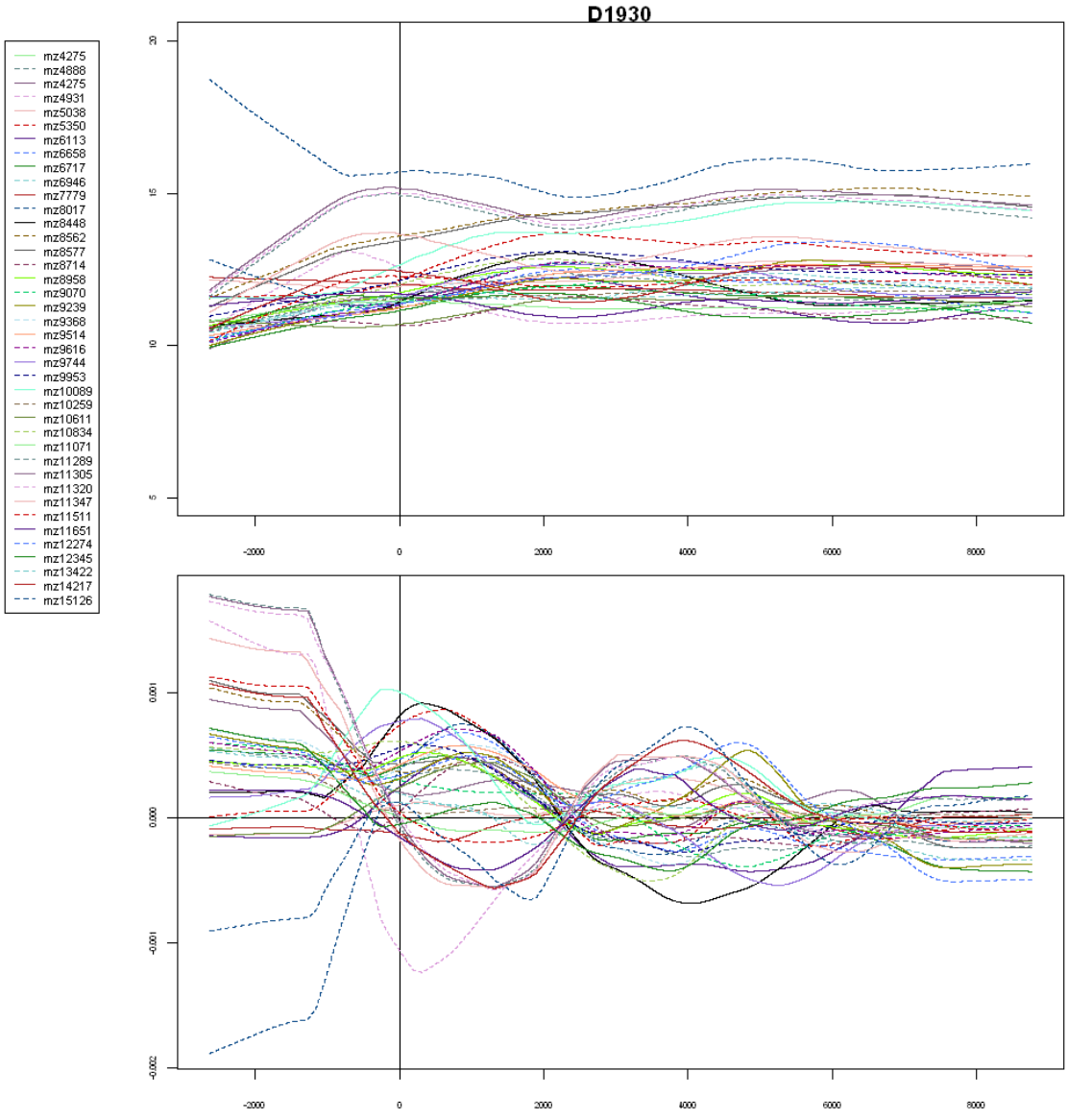

D1922
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
5
10
15
20
-6000
-4000
-2000
0
2000
4000
6000
-1e-03
-5e-04
0e+00
5e-04
1e-03

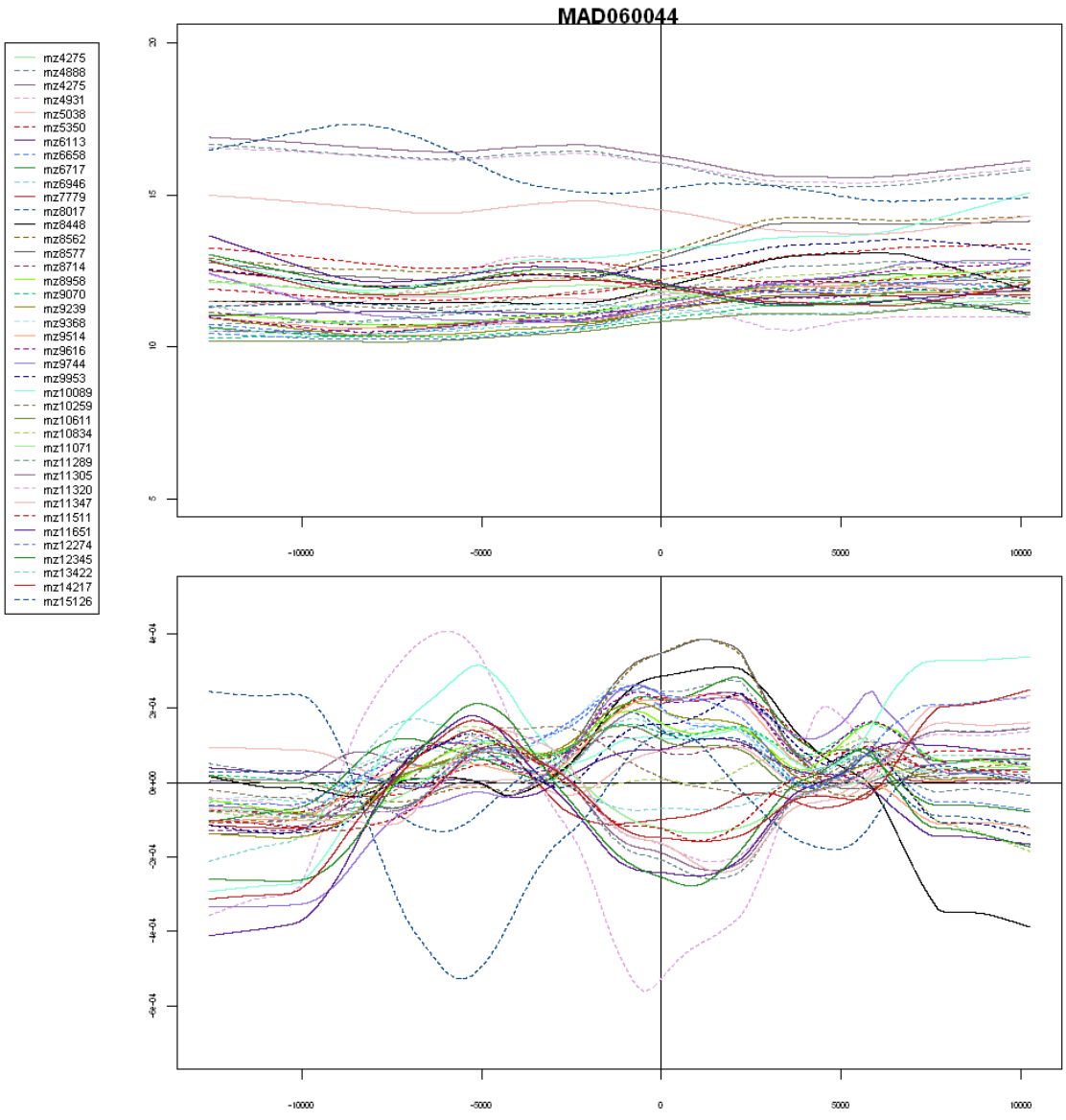

D0902
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
5
10
15
20
-6000
-4000
-2000
0
2000
4000
6000
-5e-04
0e+00
5e-04
1e-03

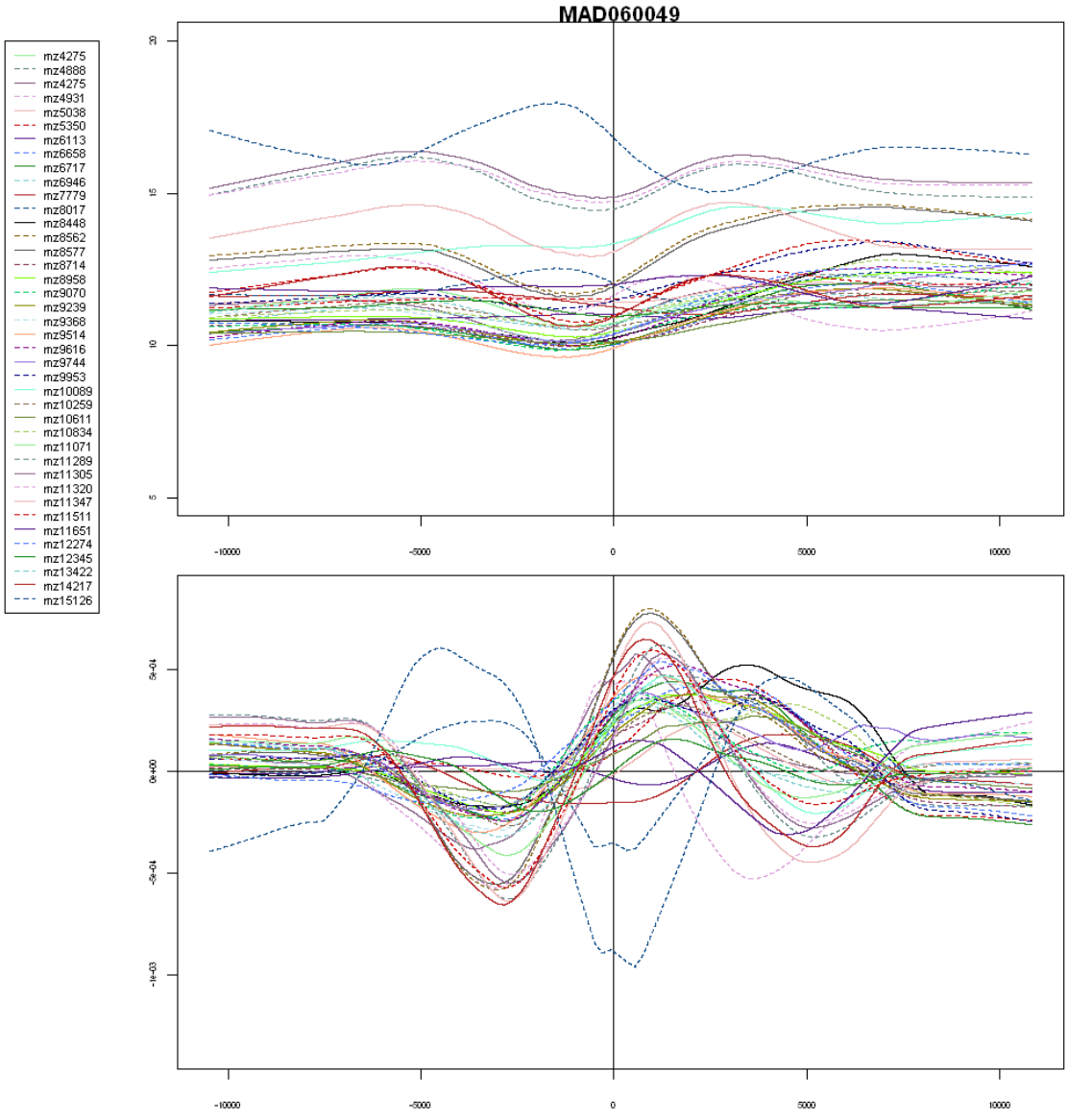

D1349
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-10000
-5000
0
5000
0.002
0.001
0.000
-0.001
-0.002
-10000
-5000
0
5000

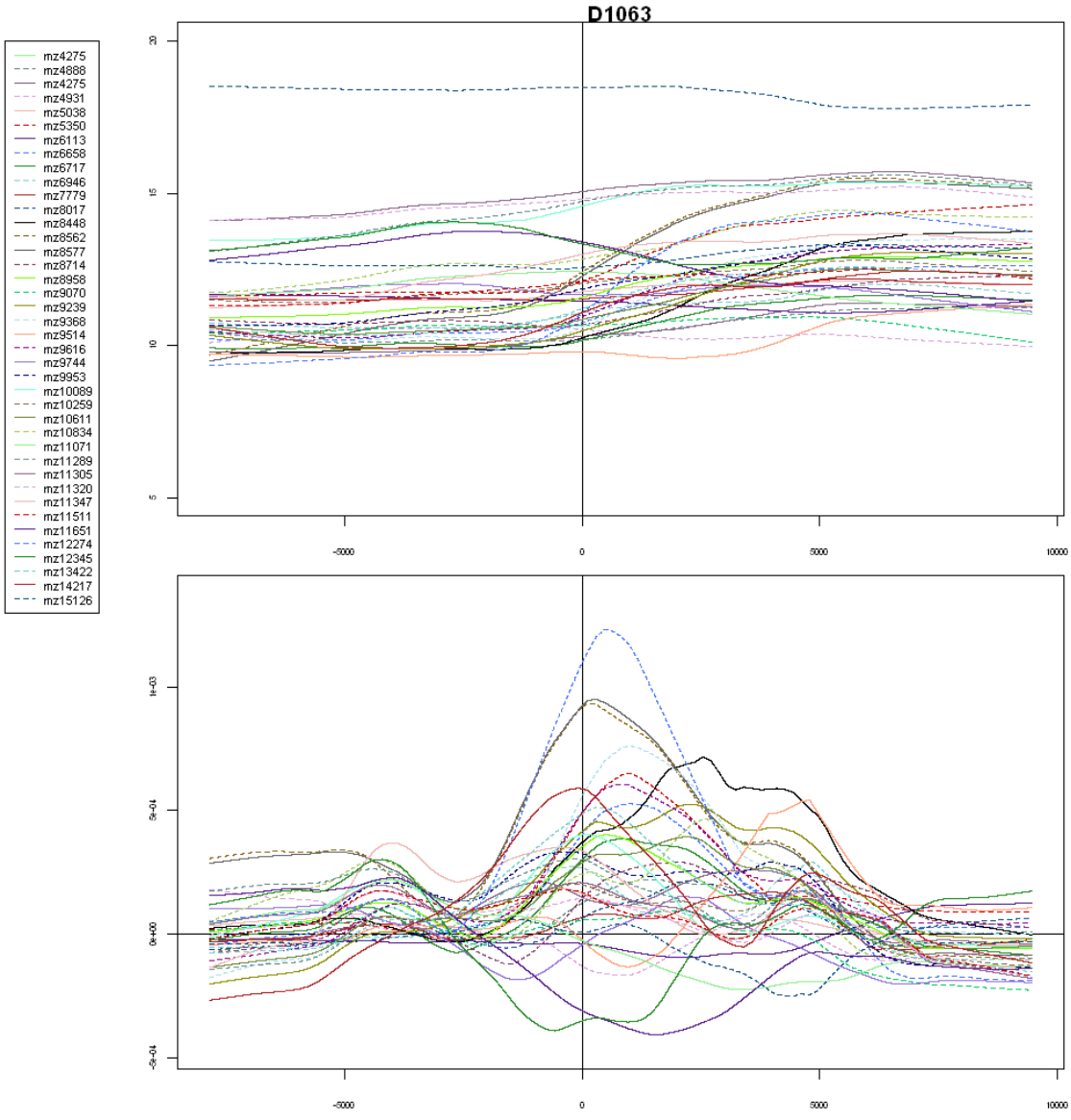

D1923
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
-5000
0
5000
10000


D1925
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-2000
0
2000
4000
6000
0.002
0.001
0.000
-0.001
-0.002
-2000
0
2000
4000
6000


D1930
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-2000
0
2000
4000
6000
8000
0.001
0.000
-0.001
-0.002
-2000
0
2000
4000
6000
8000

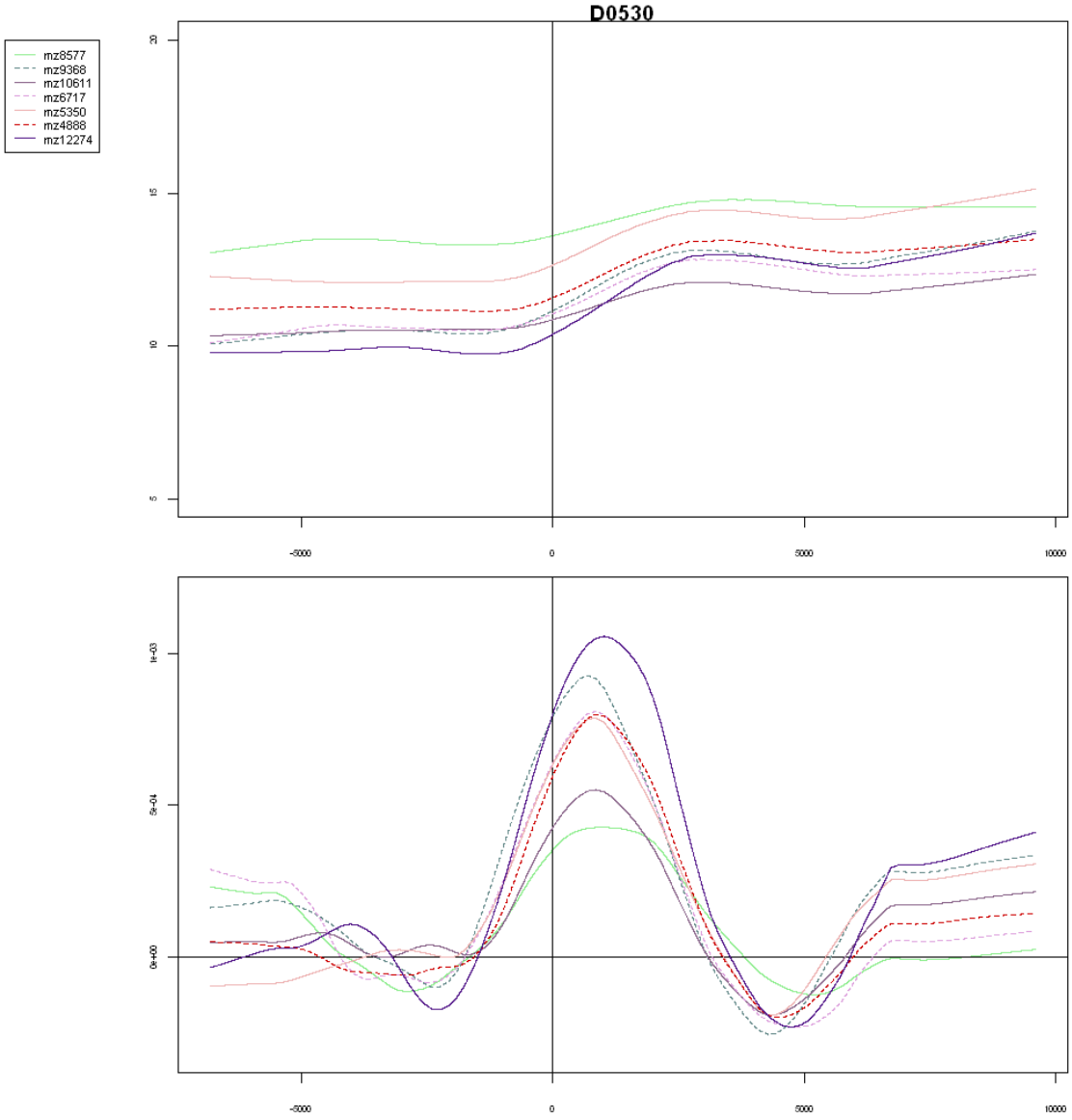

MAD060044
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-10000
-5000
0
5000
10000
4e-04
2e-04
0e+00
-2e-04
-4e-04
-6e-04
-10000
-5000
0
5000
10000

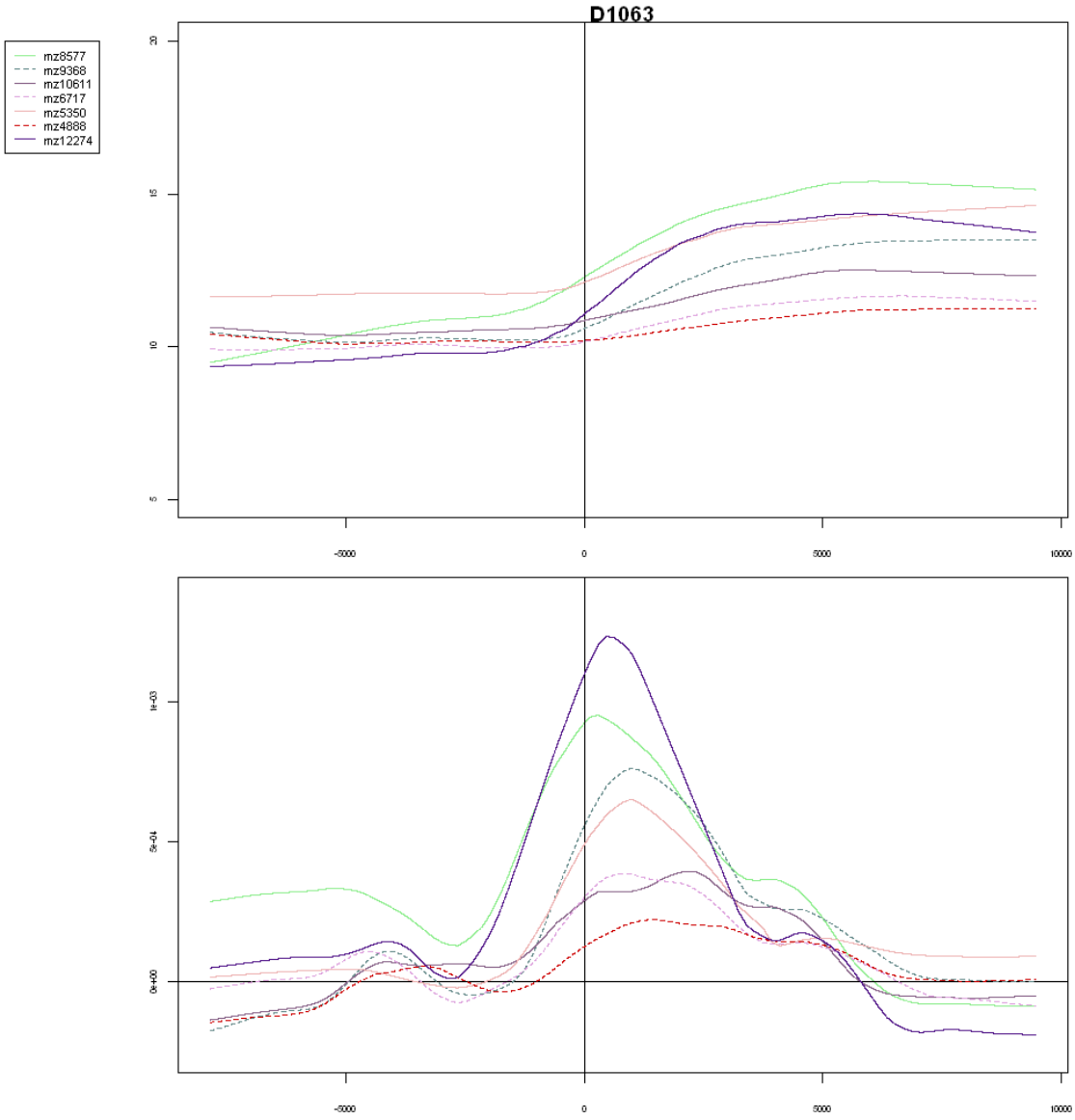

MAD060049
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-10000
-5000
0
5000
10000
5e-04
0e+00
-5e-04
-1e-03
-10000
-5000
0
5000
10000

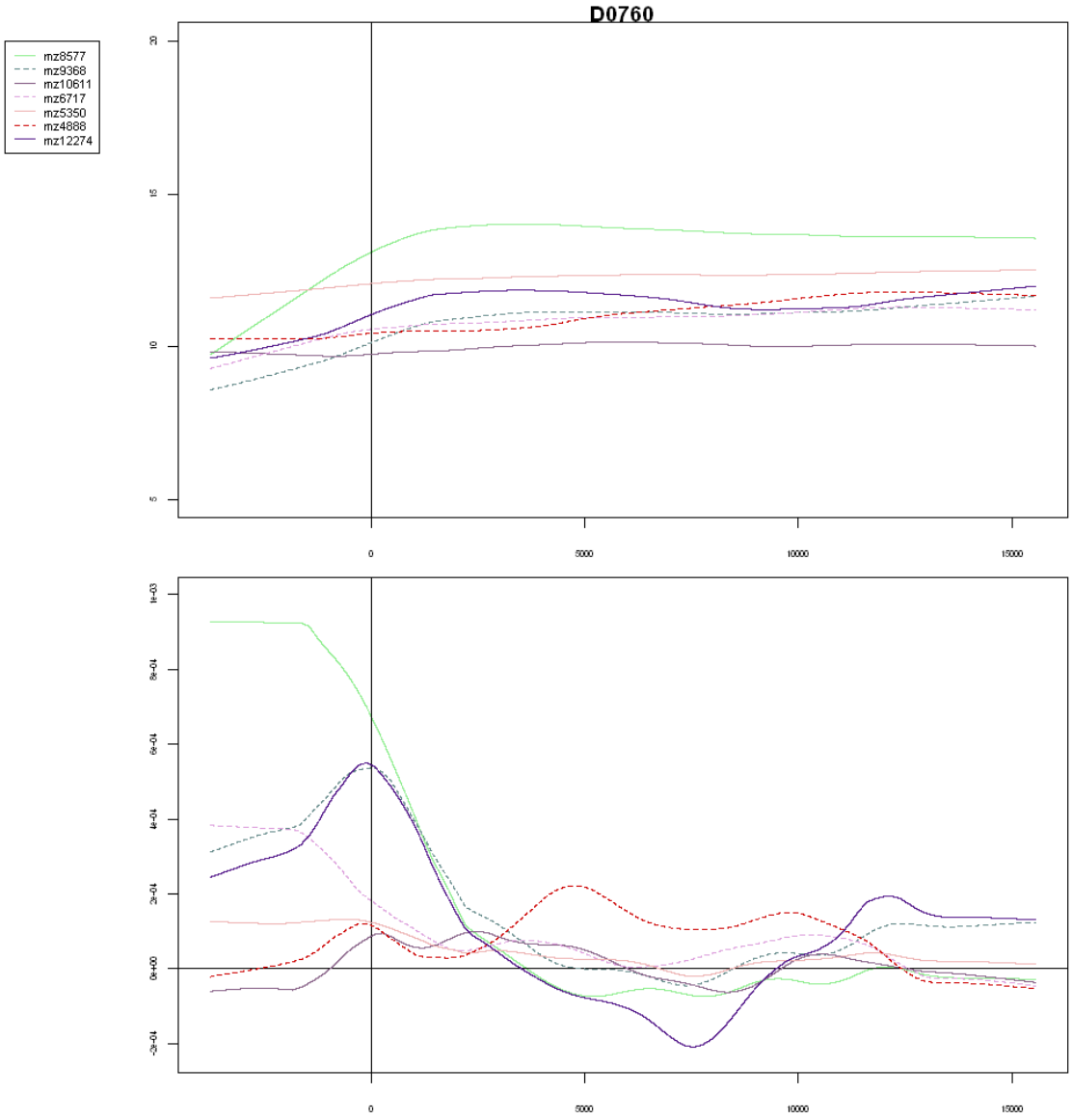

D1063
mz4275
mz4888
mz4275
mz4931
mz5038
mz5350
mz6113
mz6658
mz6717
mz6946
mz7779
mz8017
mz8448
mz8562
mz8577
mz8714
mz8958
mz9070
mz9239
mz9368
mz9514
mz9616
mz9744
mz9953
mz10089
mz10259
mz10611
mz10834
mz11071
mz11289
mz11305
mz11320
mz11347
mz11511
mz11651
mz12274
mz12345
mz13422
mz14217
mz15126
20
15
10
5
-5000
0
5000
10000
1e-03
5e-04
0e+00
-5e-04
-5000
0
5000
10000

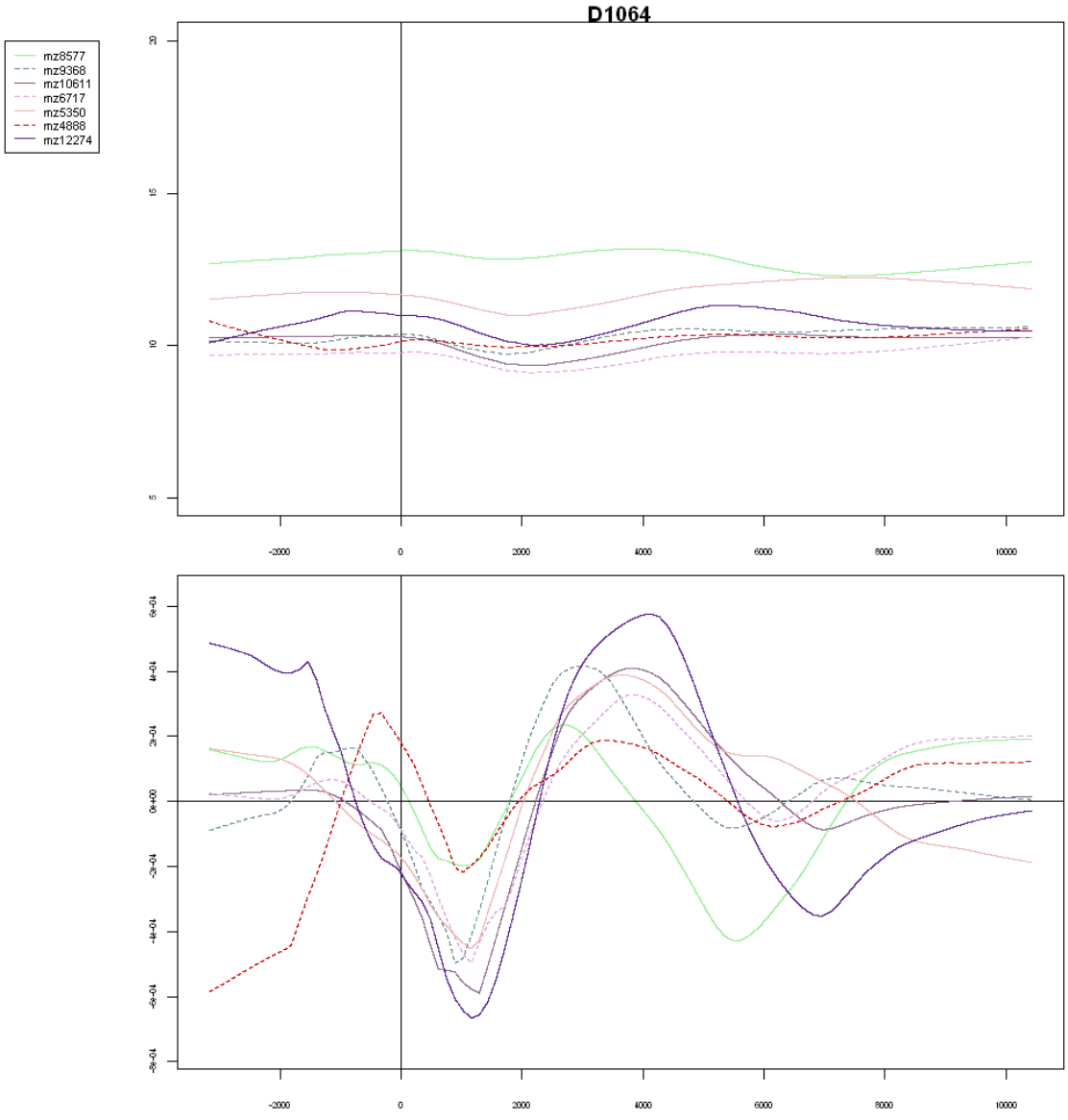
Appendix D

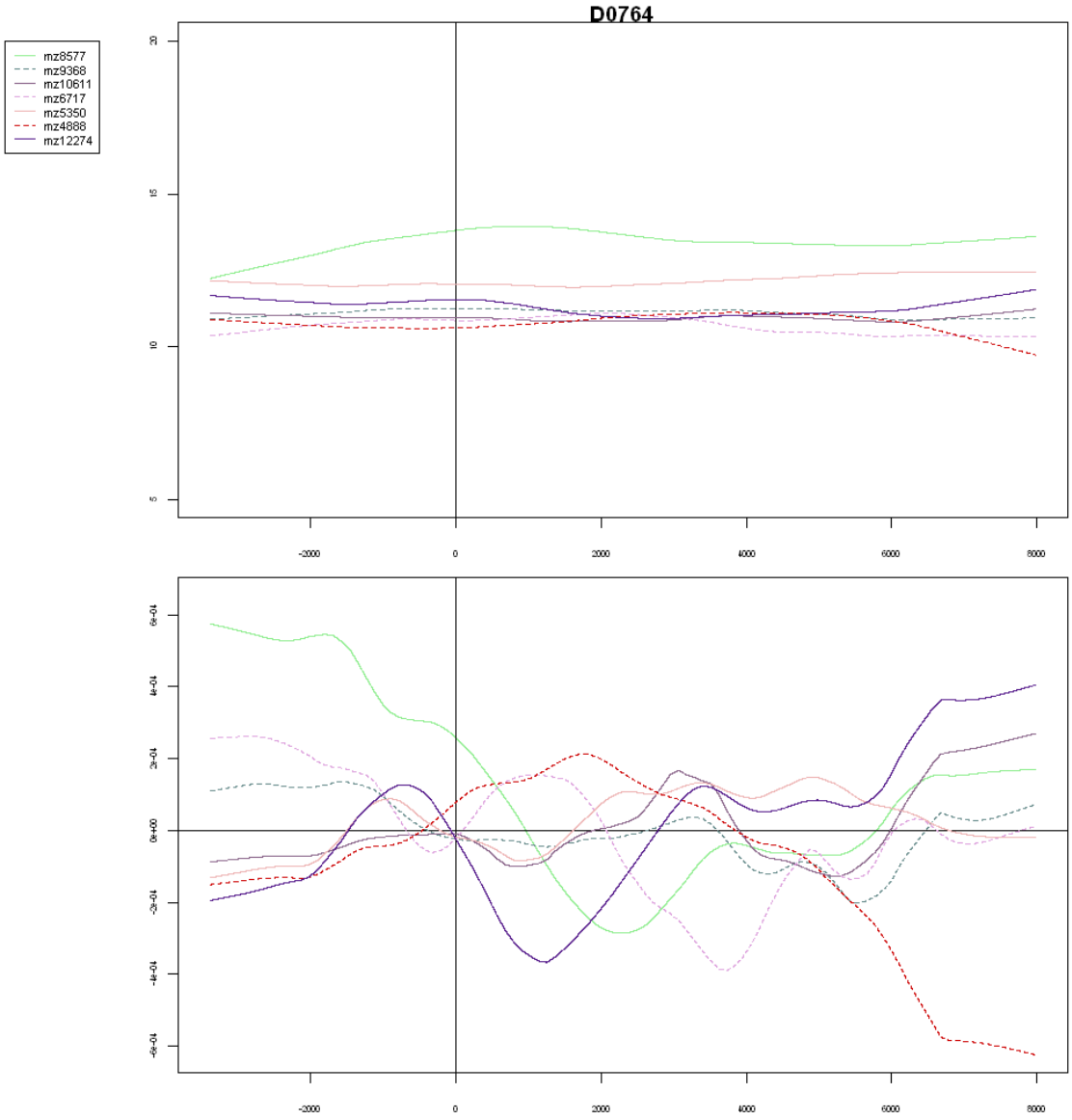
Note: Refer to Figure 32 for graph and axes information pertaining to all figures in Appendix D.

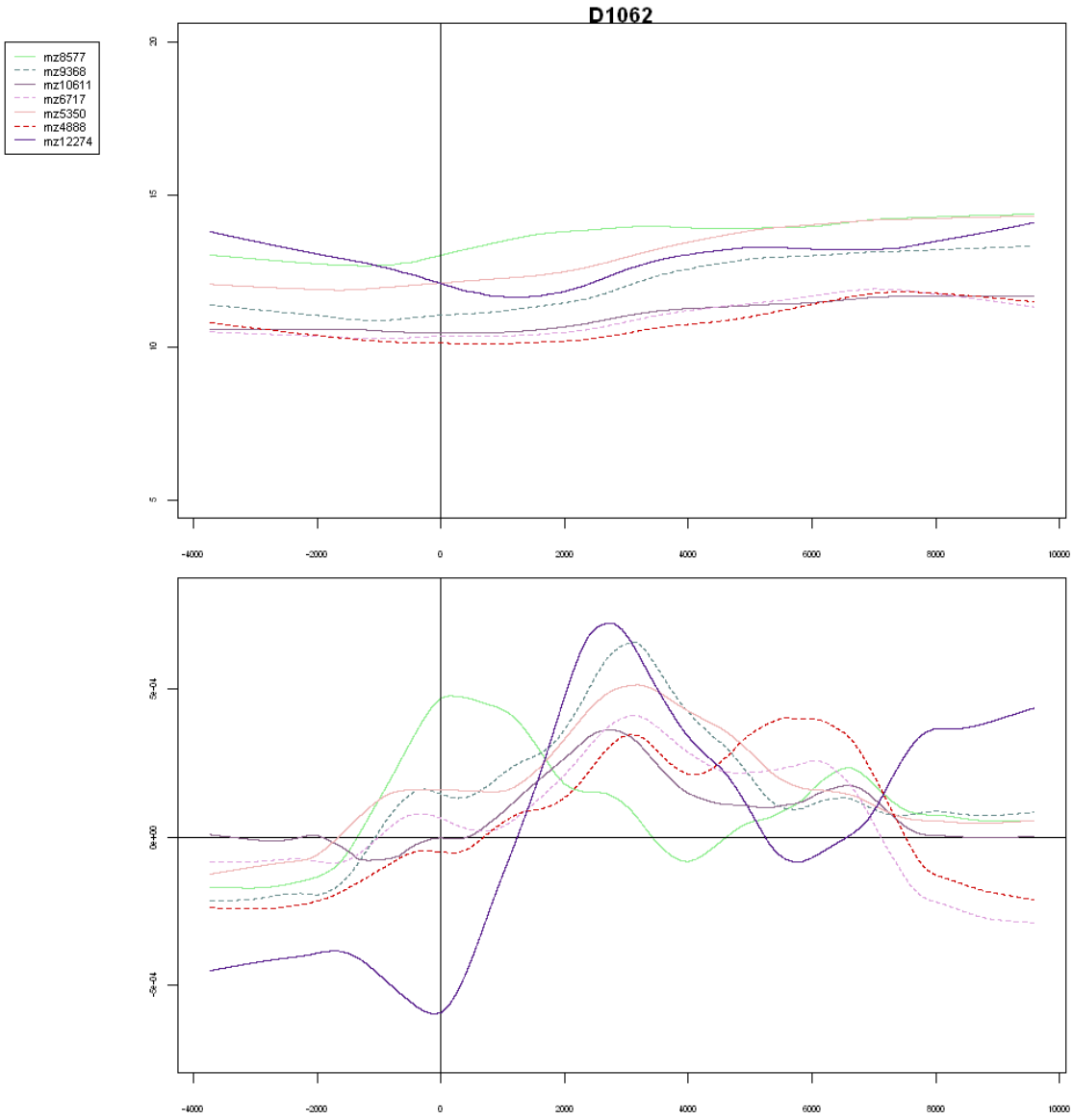


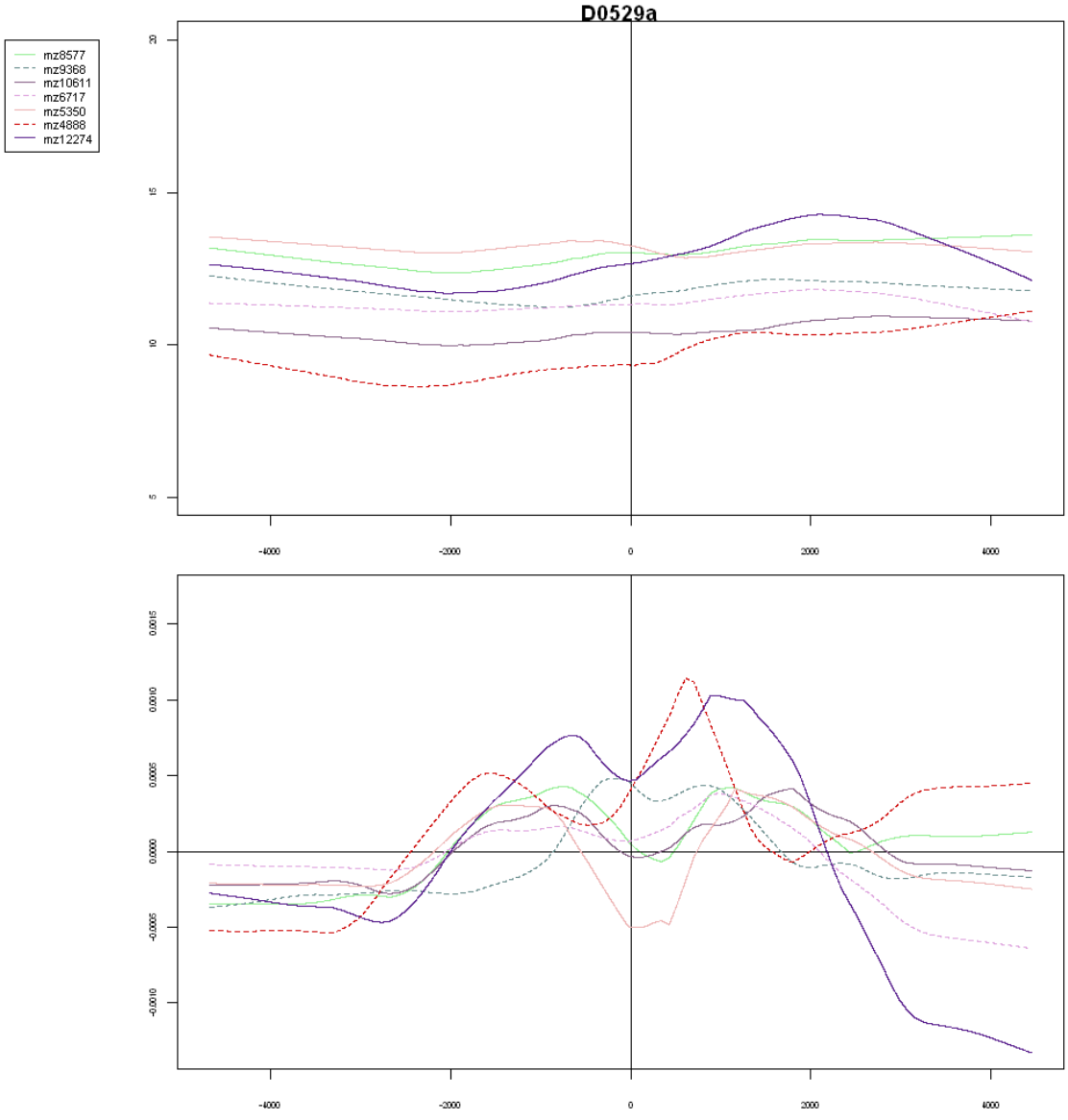

D1063
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
-5000
0
5000
10000

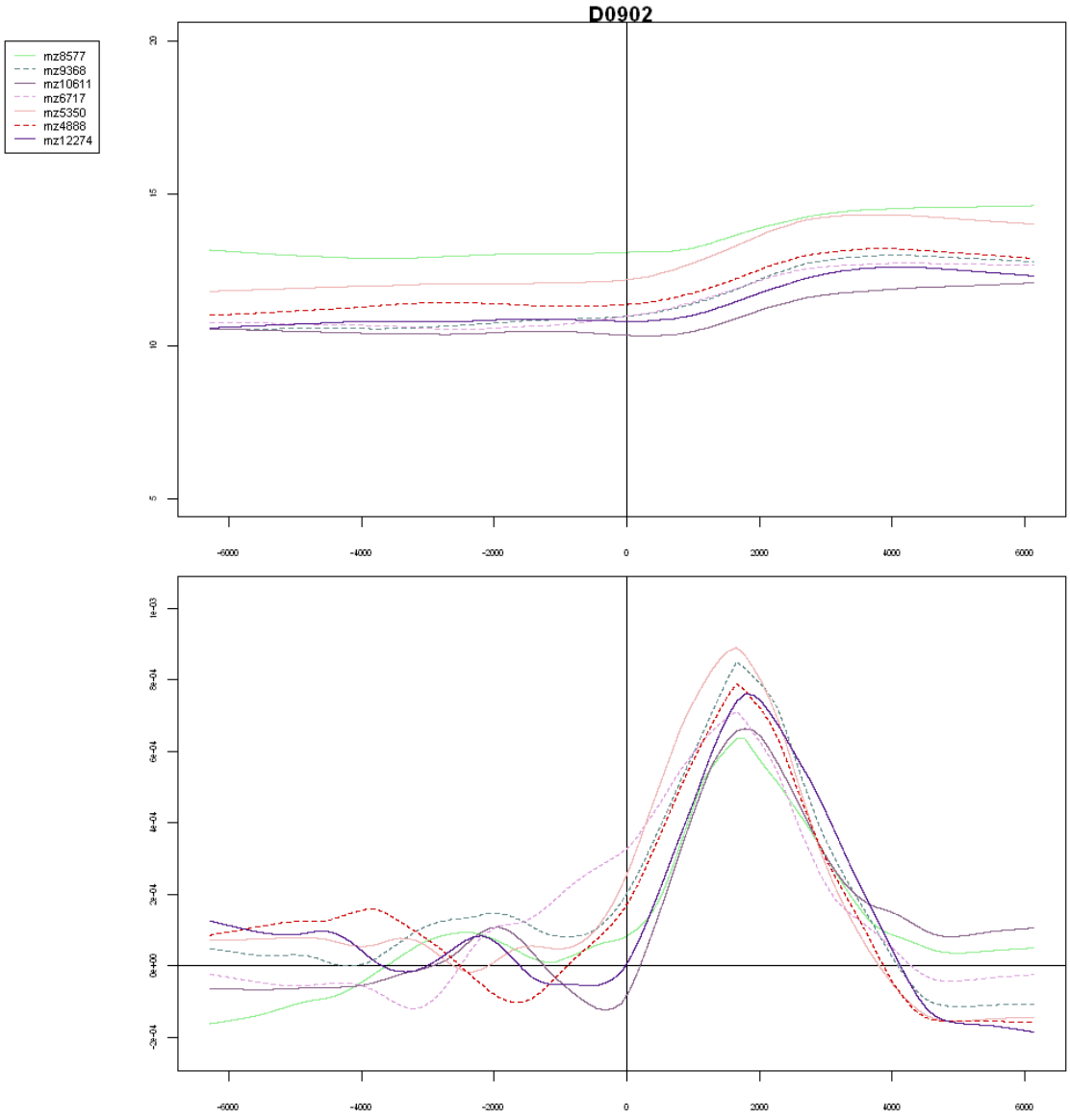

D0760
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
0
5000
10000
15000
1e-03
8e-04
6e-04
4e-04
2e-04
0e+00
-2e-04

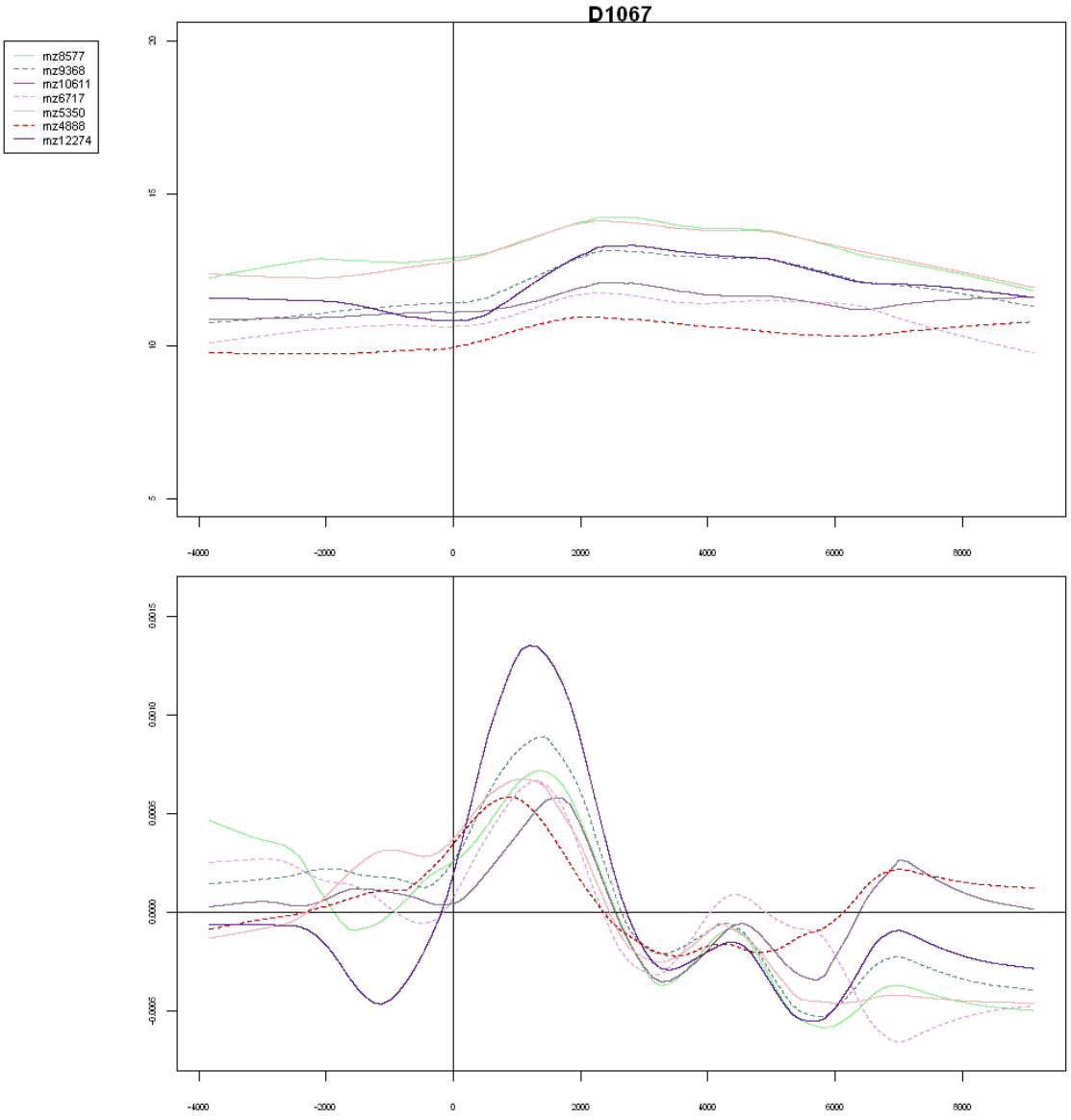

D1064
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-2000
0
2000
4000
6000
8000
10000
6e-04
4e-04
2e-04
0e+00
-2e-04
-4e-04
-6e-04
-8e-04

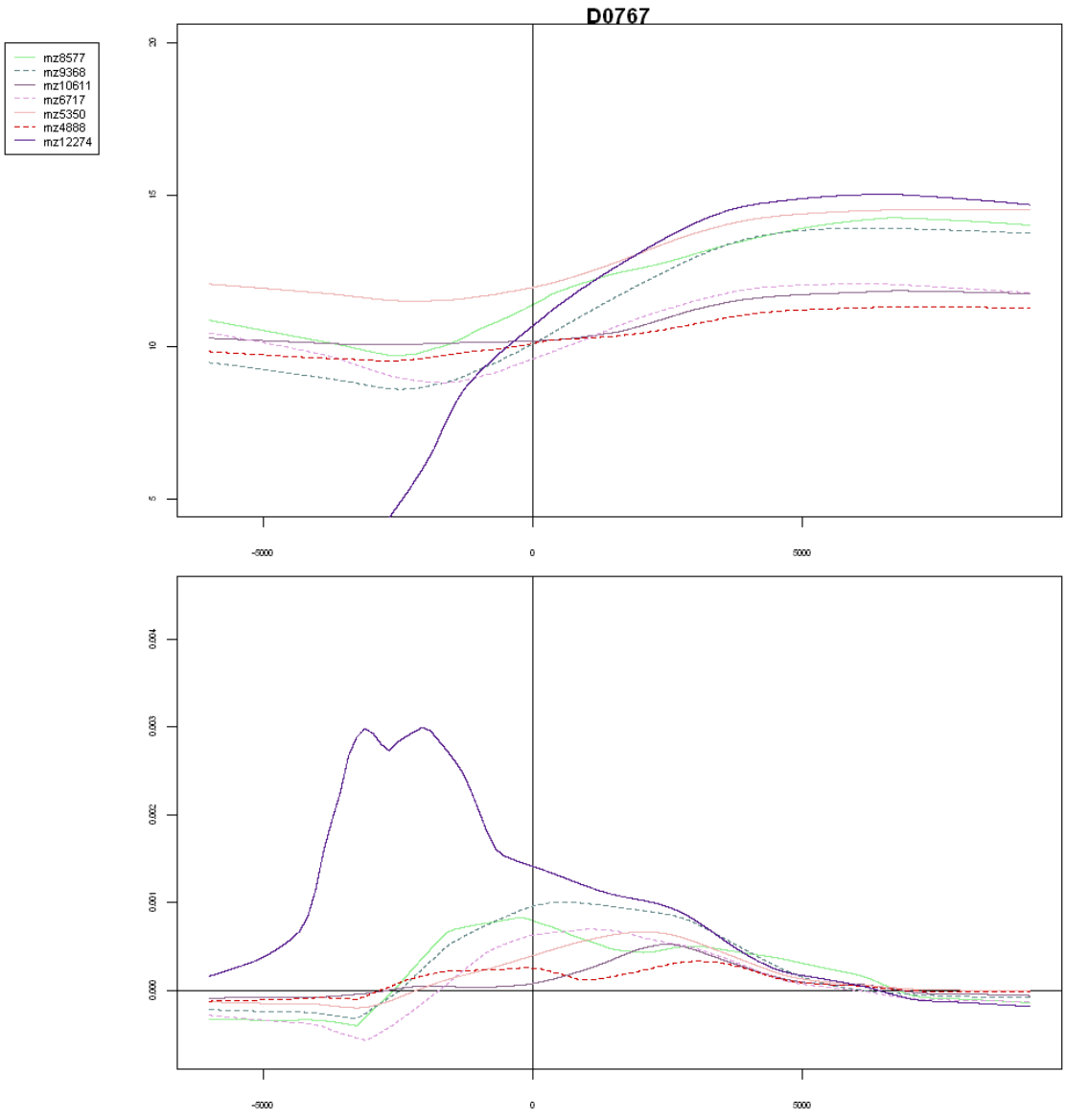

D0764
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-2000
0
2000
4000
6000
8000
6e-04
4e-04
2e-04
0e+00
-2e-04
-4e-04
-6e-04

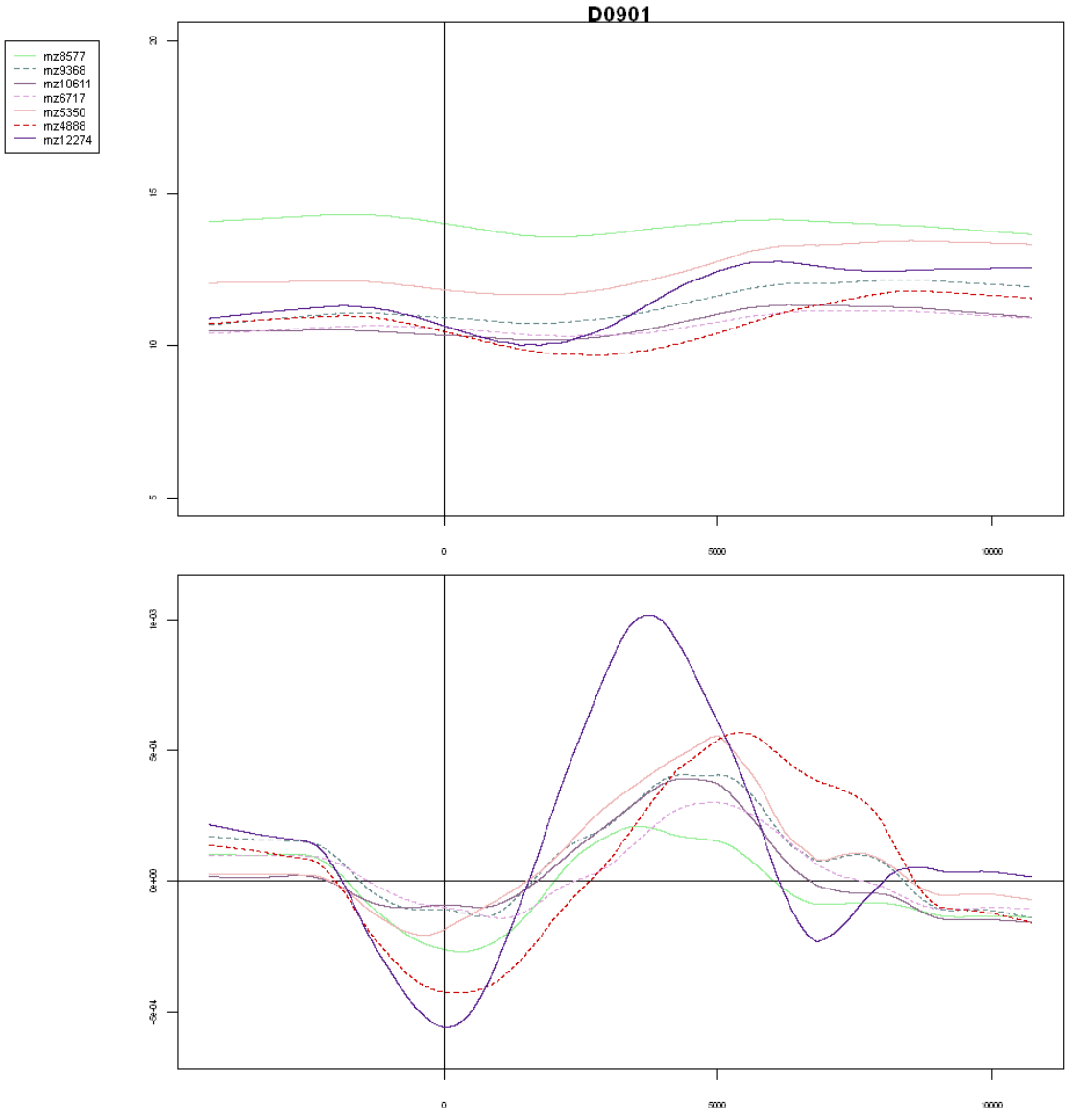

D1062
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-4000
-2000
0
2000
4000
6000
8000
10000
5e-04
0e+00
-5e-04

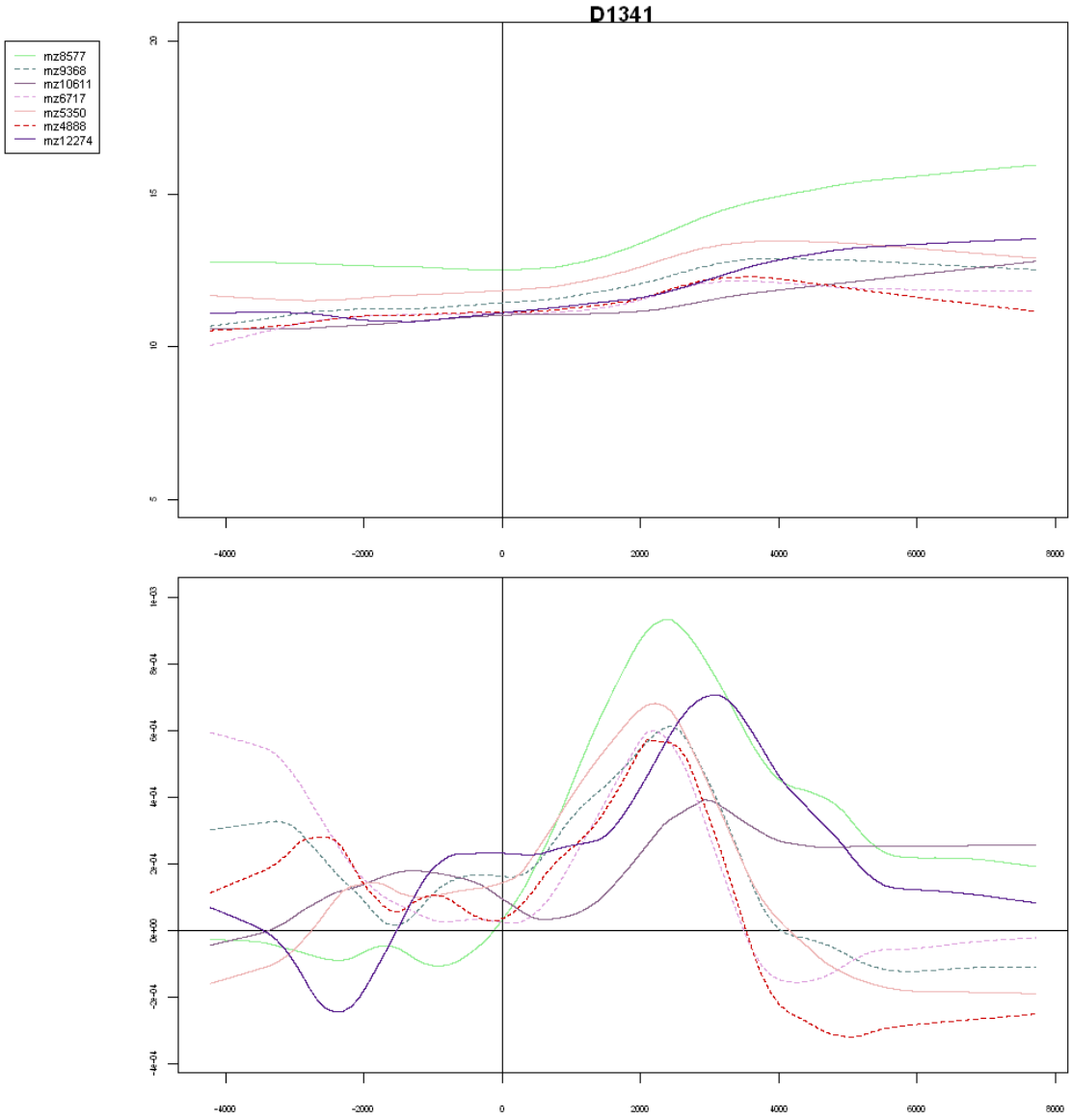

D0529a
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-4000
-2000
0
2000
4000
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-4000
-2000
0
2000
4000

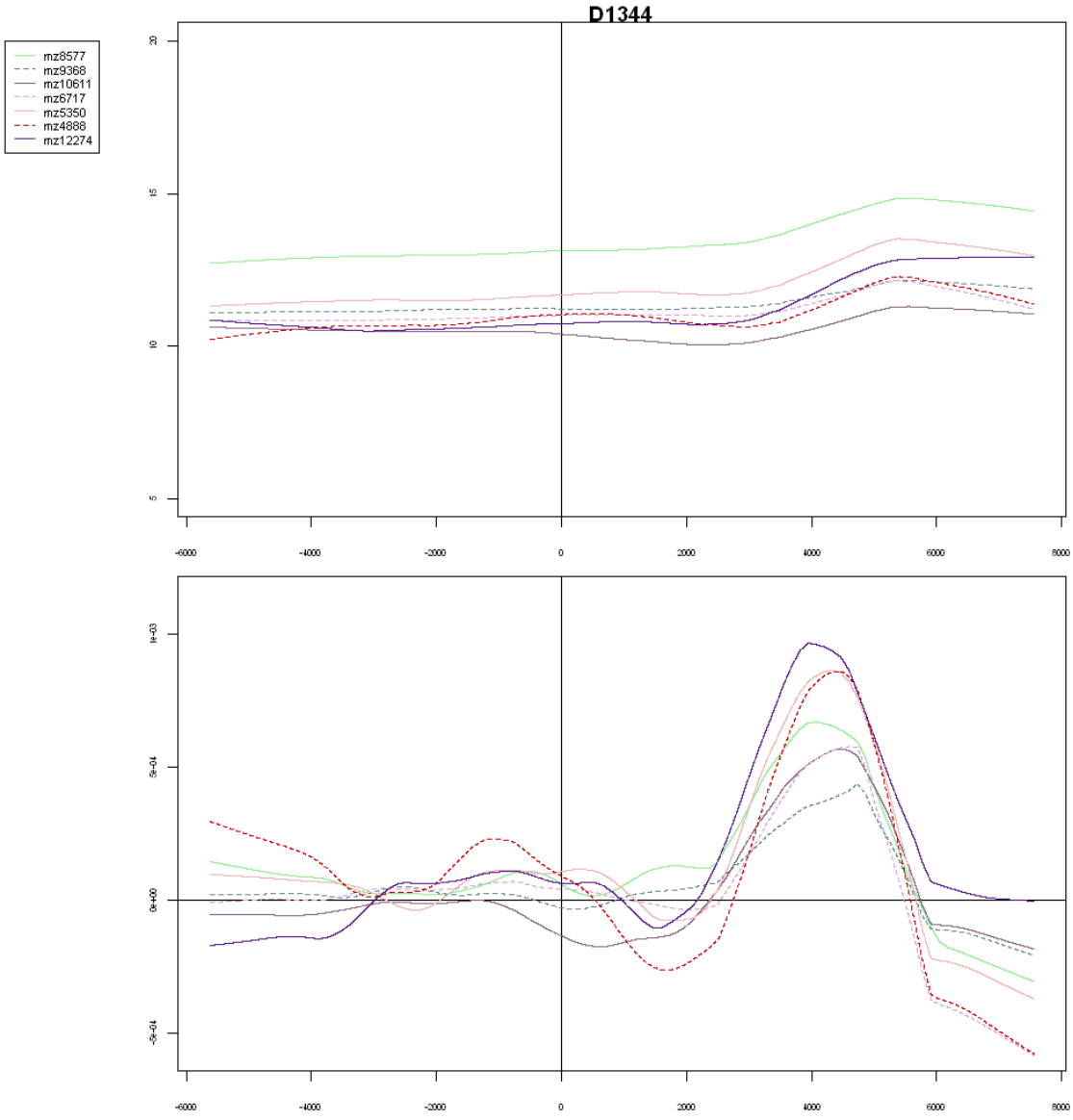

D0902
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-6000
-4000
-2000
0
2000
4000
6000
1e-03
8e-04
6e-04
4e-04
2e-04
0e+00
-2e-04

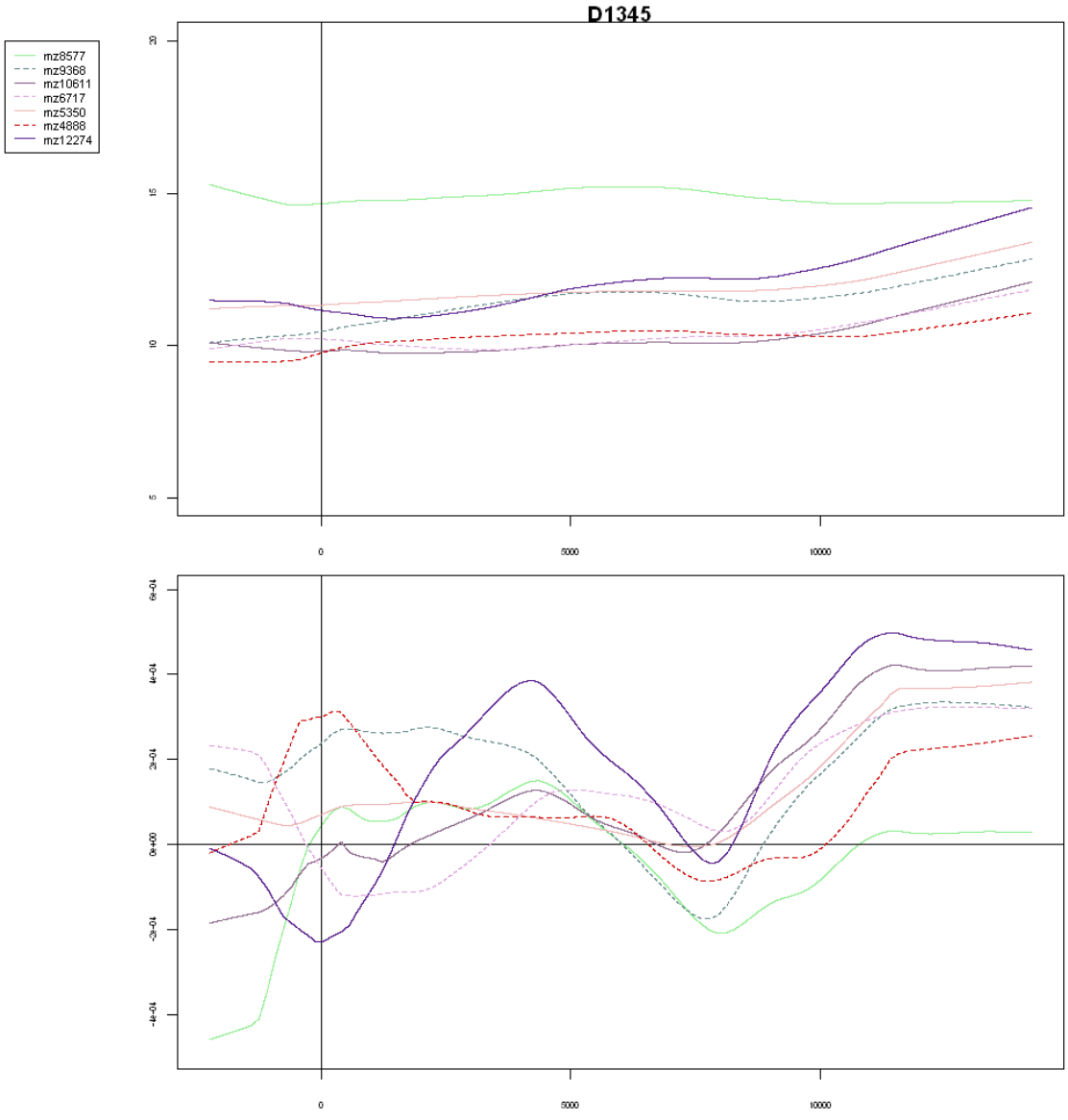

D1067
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-4000
-2000
0
2000
4000
6000
8000
0.0015
0.0010
0.0005
0.0000
-0.0005

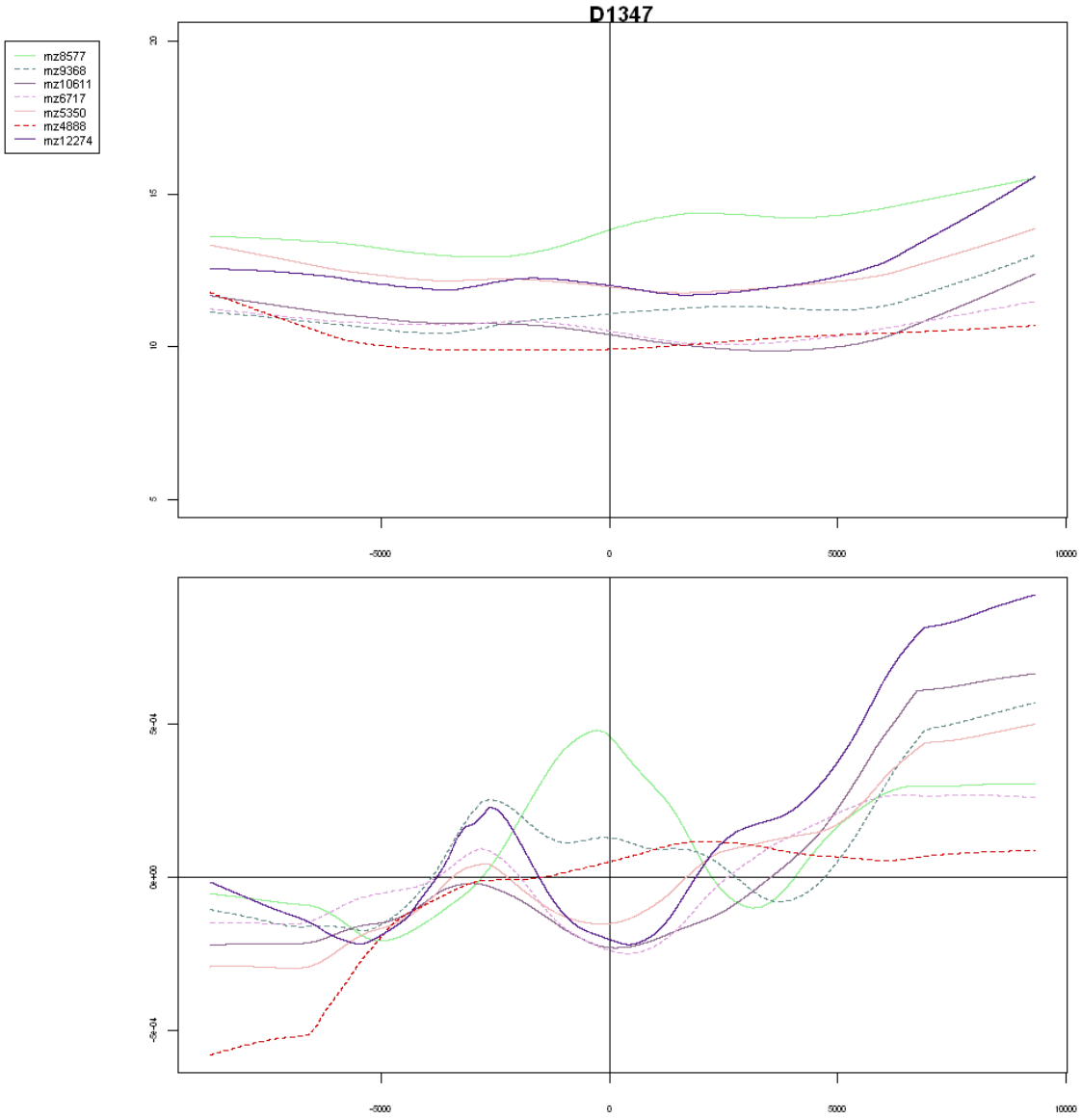

D0767
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-5000
0
5000
0.004
0.003
0.002
0.001
0.000
-5000
0
5000

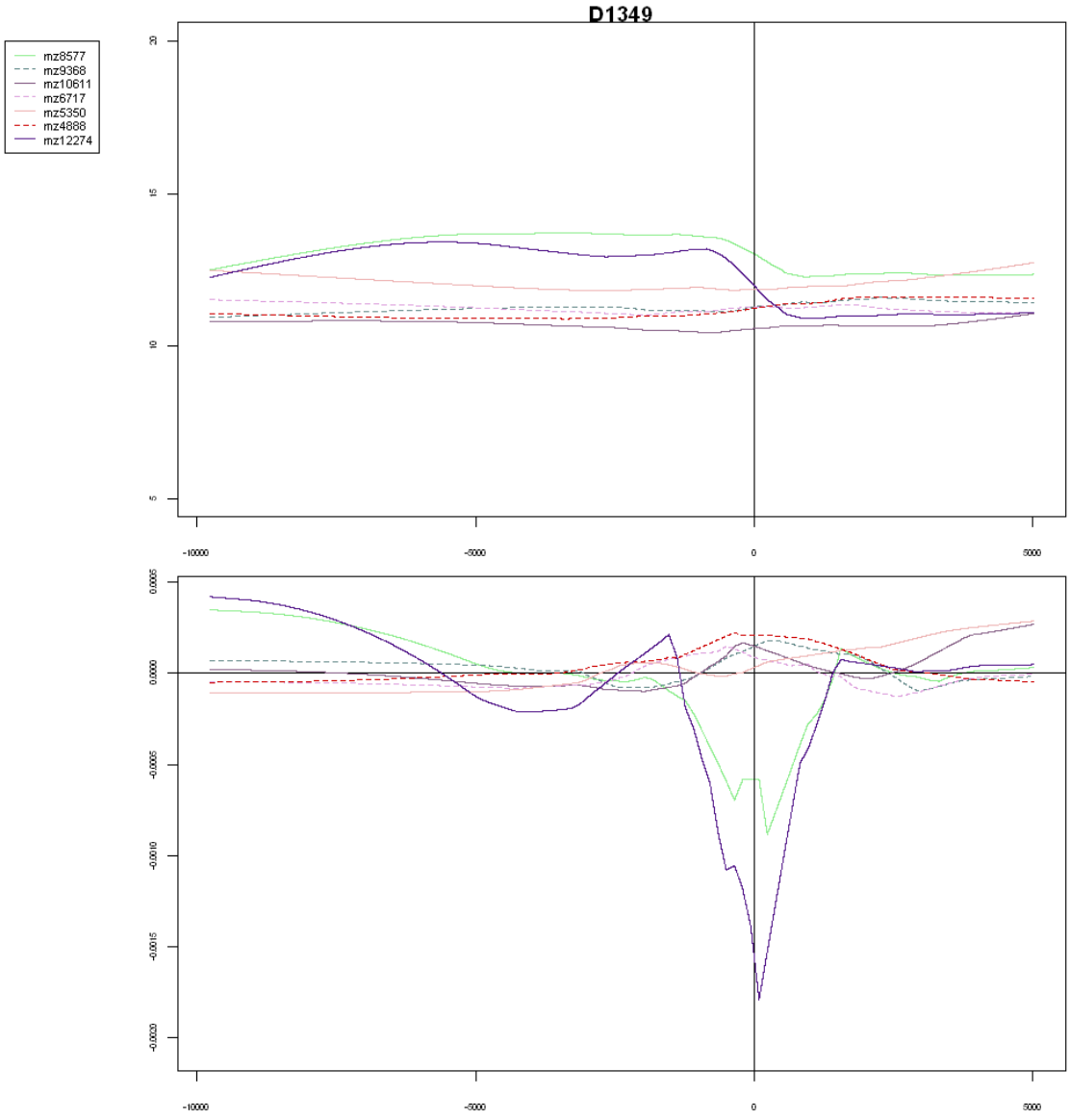

D0901
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
15
10
5
0
5000
10000
1e+03
5e+04
0e+00
-5e+04
0
5000
10000

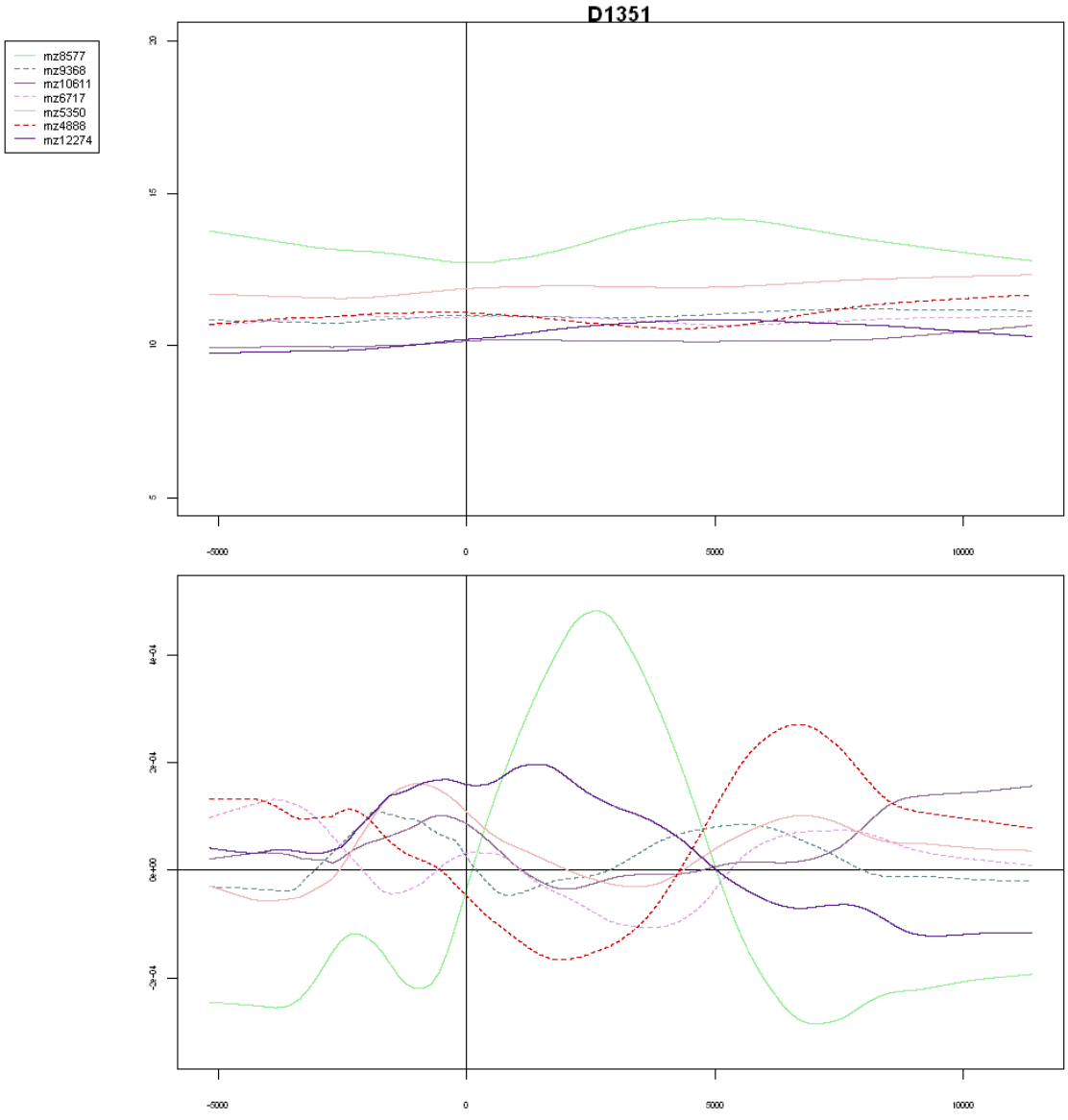

D1341
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-4000
-2000
0
2000
4000
6000
8000
1e-03
8e-04
6e-04
4e-04
2e-04
0e+00
-2e-04
-4e-04

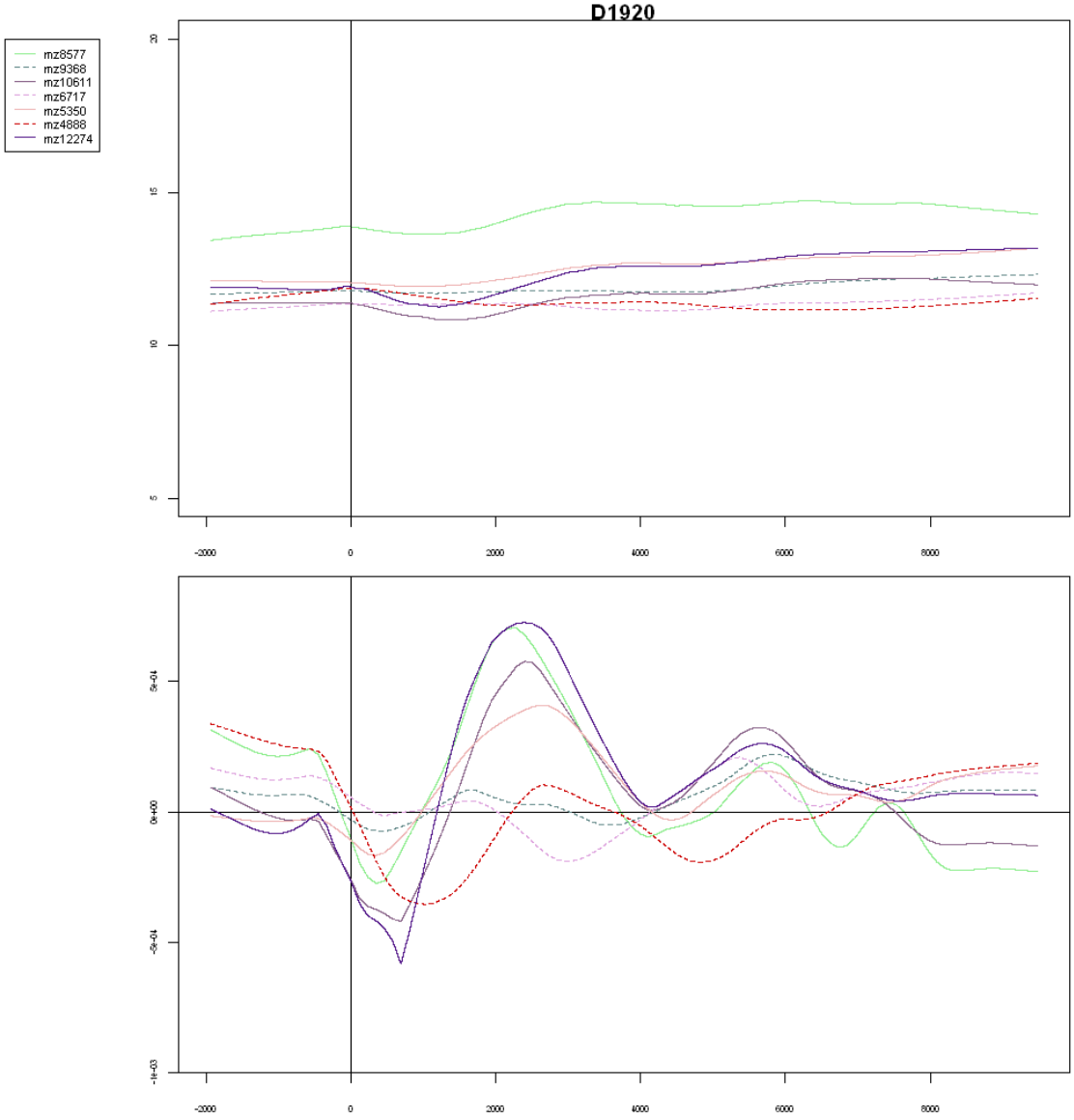

D1344
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-6000
-4000
-2000
0
2000
4000
6000
8000
1e+03
5e+04
0e+00
-5e+04

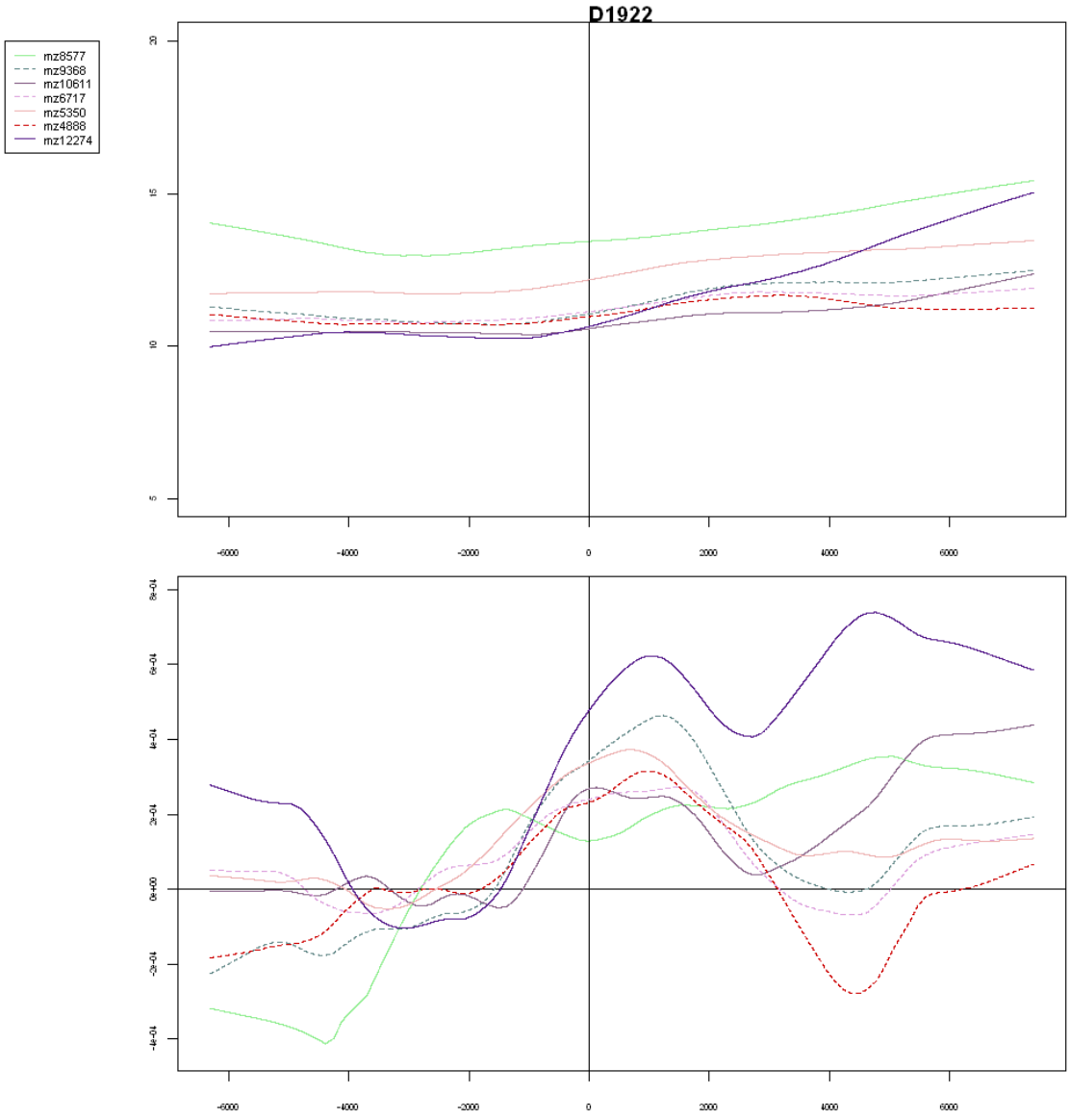

D1345
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
0
5000
10000
6e-04
4e-04
2e-04
0e+00
-2e-04
-4e-04
0
5000
10000

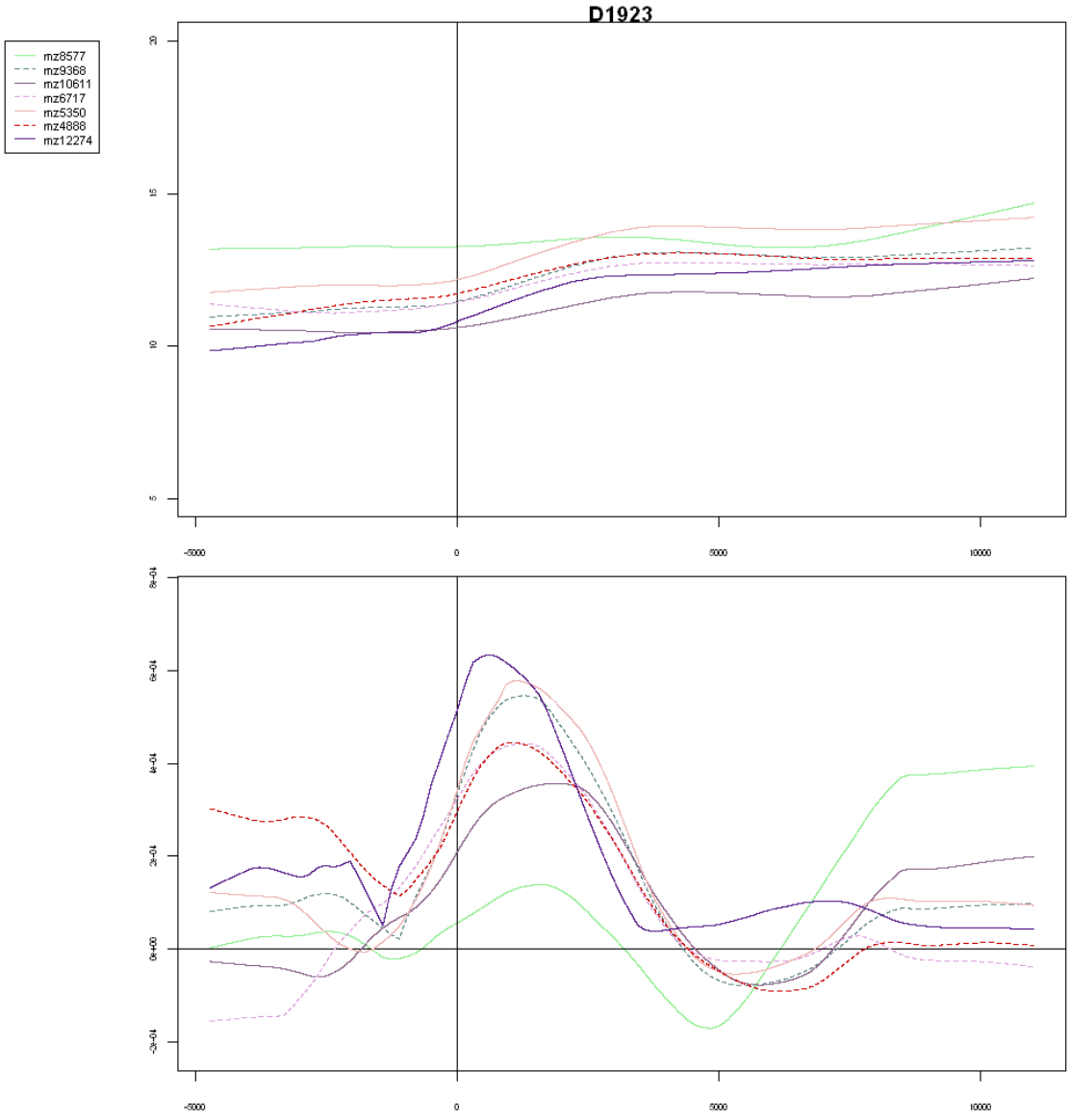

D1347
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-5000
0
5000
10000
5e-04
0e+00
-5e-04
-5000
0
5000
10000

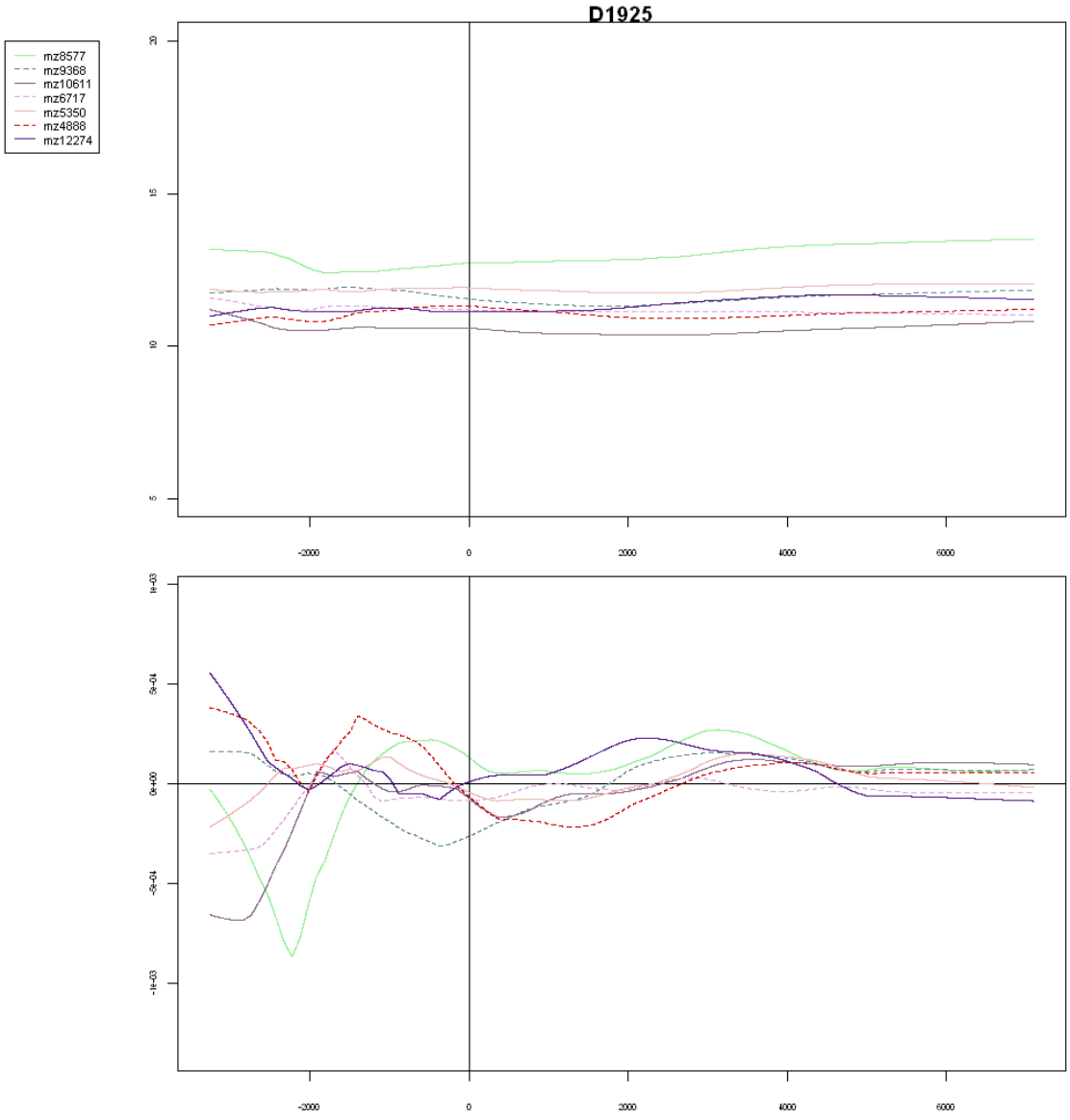

D1349
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-10000
-5000
0
5000
0.0005
0.0000
-0.0005
-0.0010
-0.0015
-0.0020
-10000
-5000
0
5000

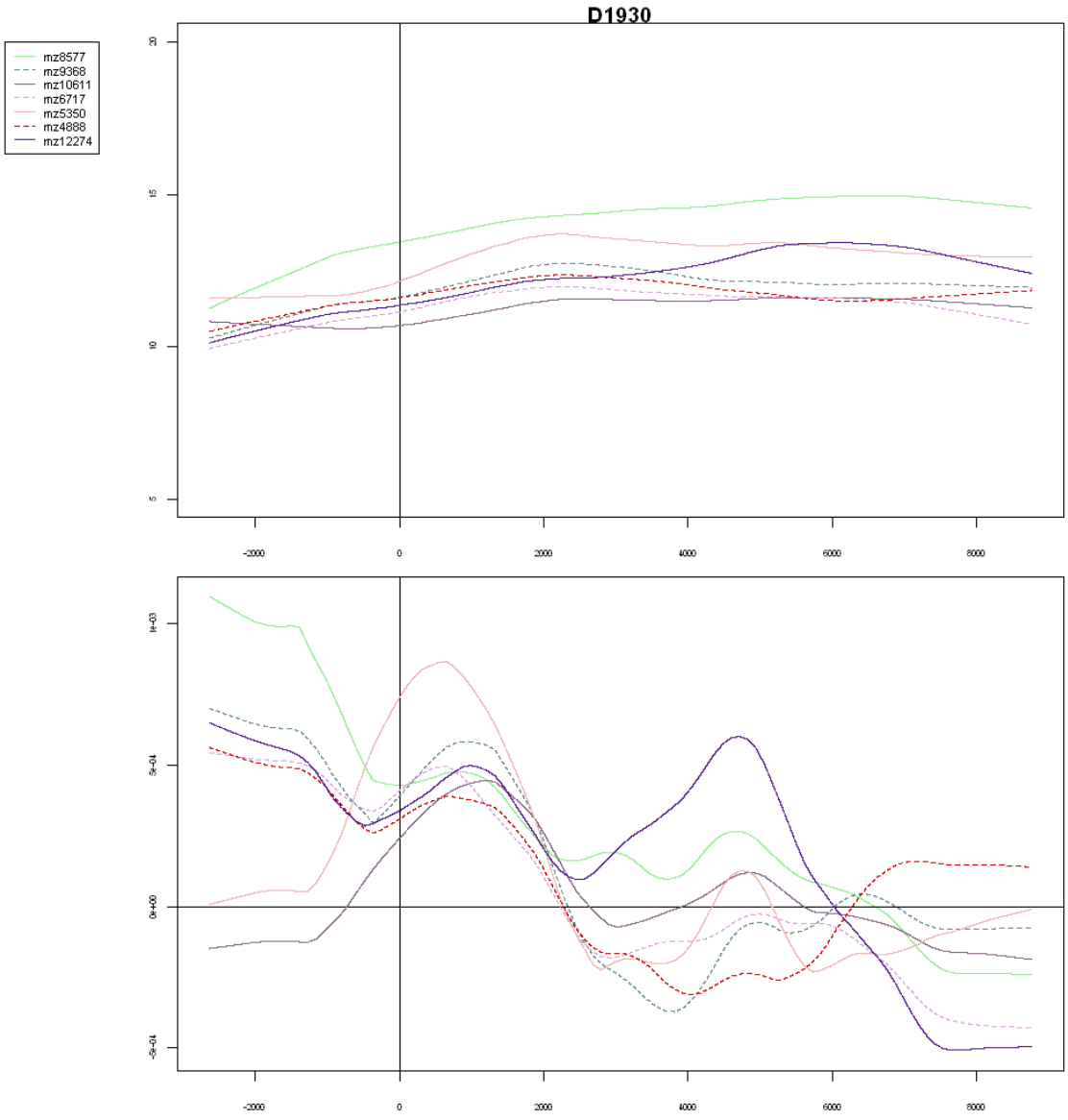

D1351
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-5000
0
5000
10000
4e-04
2e-04
0e+00
-2e-04
-5000
0
5000
10000

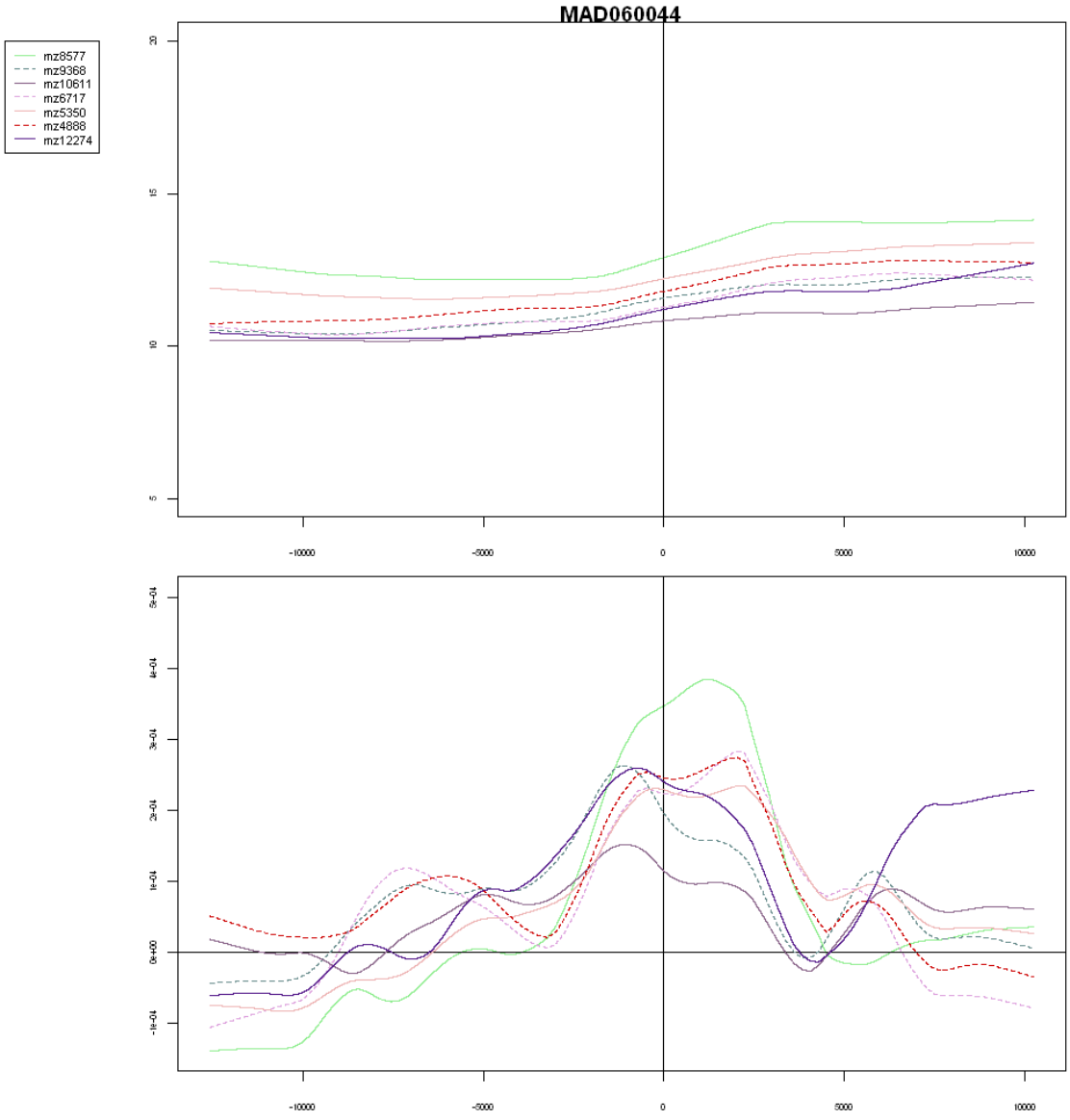

D1920
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-2000
0
2000
4000
6000
8000
5e-04
0e+00
-5e-04
-1e-03

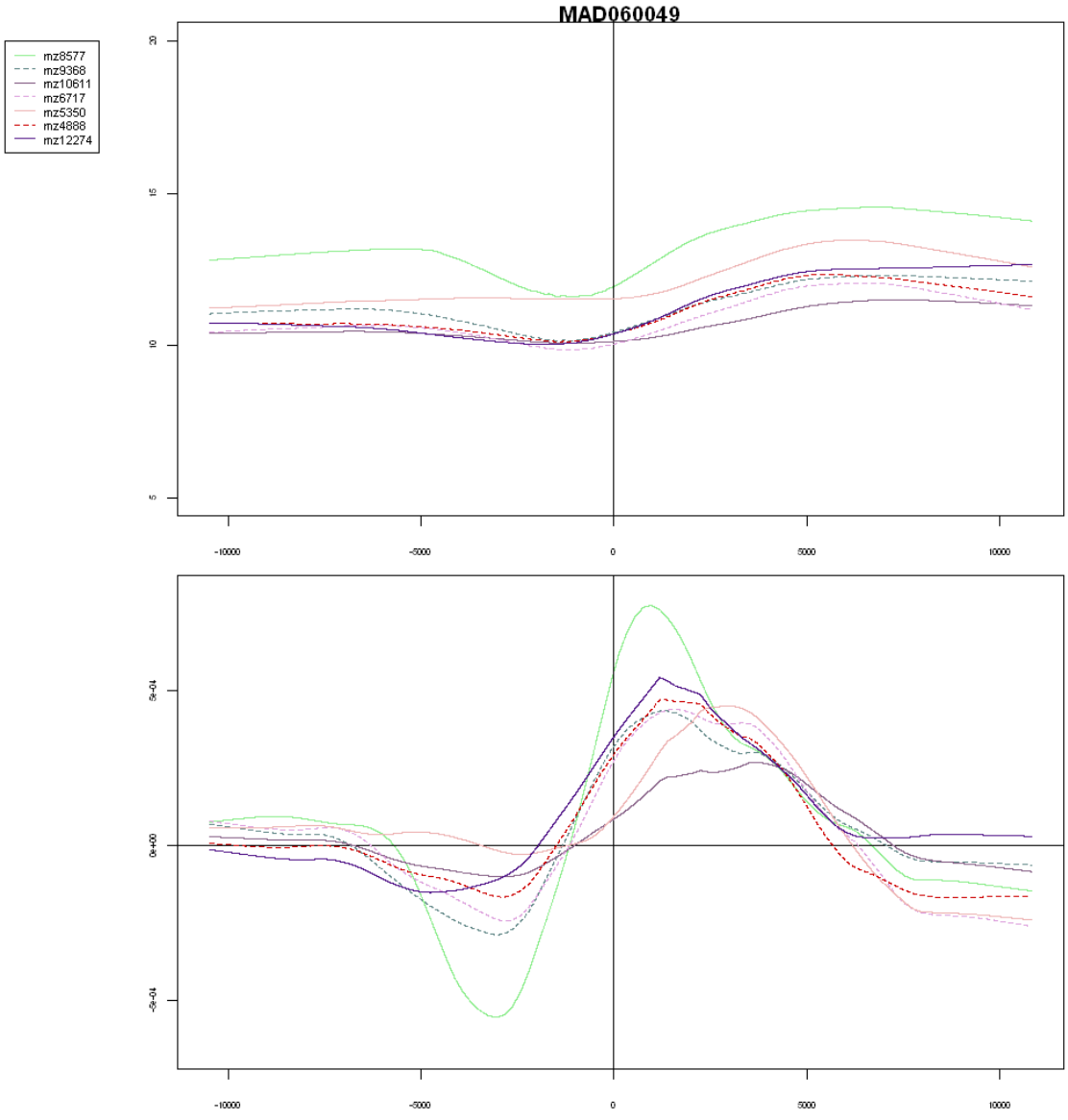

D1922
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-6000
-4000
-2000
0
2000
4000
6000
8e-04
6e-04
4e-04
2e-04
0e+00
-2e-04
-4e-04

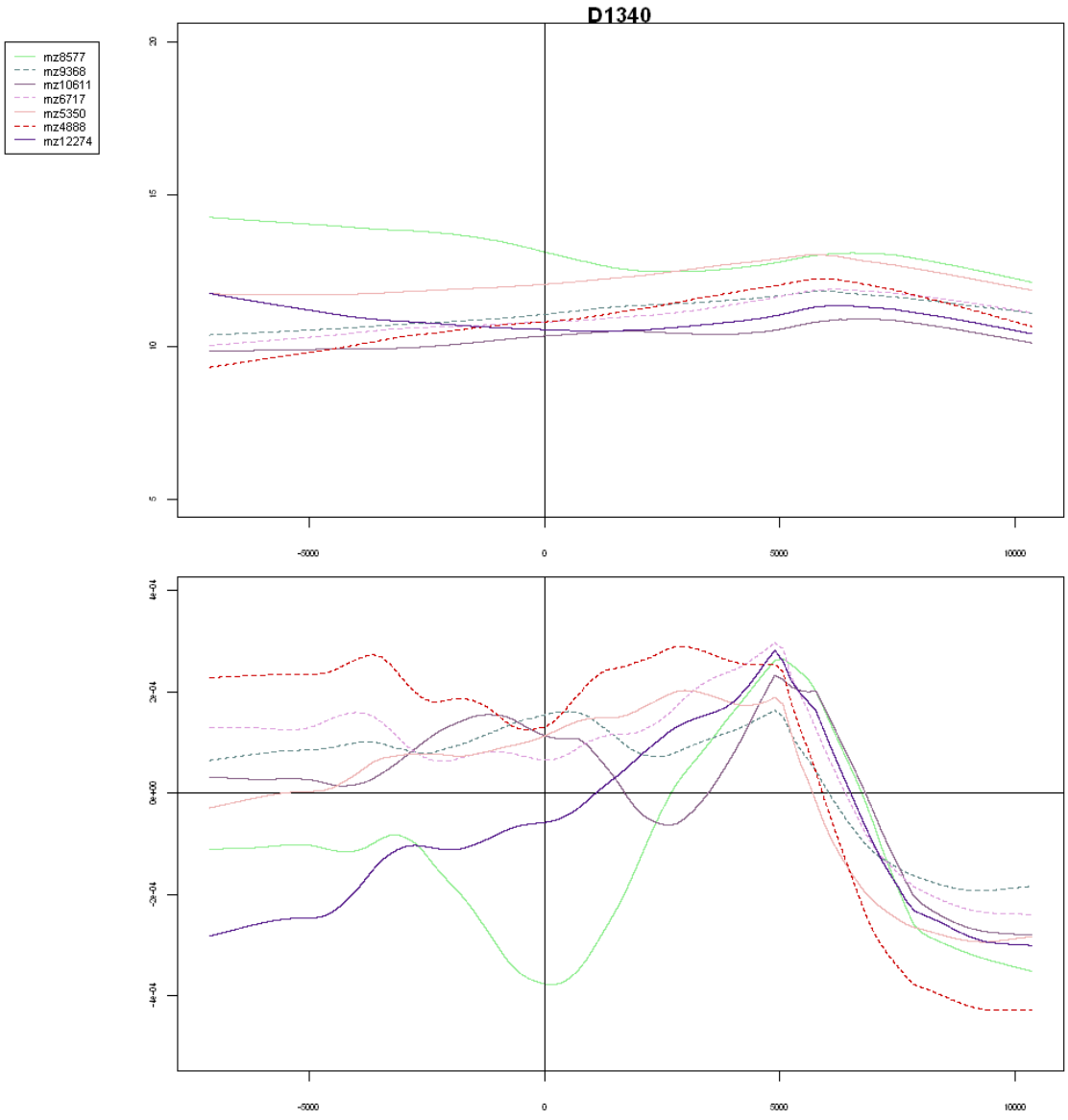

D1923
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-5000
0
5000
10000
8e-04
6e-04
4e-04
2e-04
0e+00
-2e-04


D1925
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-2000
0
2000
4000
6000
1e-03
5e-04
0e+00
-5e-04
-1e-03


D1930
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-2000
0
2000
4000
6000
8000
1e-03
5e-04
0e+00
-5e-04


MAD060044
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-10000
-5000
0
5000
10000
5e-04
4e-04
3e-04
2e-04
1e-04
0e+00
-1e-04
-10000
-5000
0
5000
10000


MAD060049
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-10000
-5000
0
5000
10000
5e-04
0e+00
-5e-04
-10000
-5000
0
5000
10000


D1340
mz8577
mz9368
mz10611
mz6717
mz5350
mz4888
mz12274
20
15
10
5
-5000
0
5000
10000
4e-04
2e-04
0e+00
-2e-04
-4e-04
-5000
0
5000
10000

REFERENCES

1. Caprioli, R. M.; Schwartz, S. A.; Chaurand, P., Profiling and Imaging Proteins in Tissue Sections by MS. *Analytical Chemistry* **2004,** 76, (5), 86A-93A.

2. Grignon, D. J.; Che, M., Clear cell renal cell carcinoma. *Clin Lab Med* **2005,** 25, (2), 305-16.

3. Caldwell, R. L.; Gonzalez, A.; Oppenheimer, S. R.; Schwartz, H. S.; Caprioli, R. M., Molecular Assessment of the Tumor Protein Microenvironment Using Imaging Mass Spectrometry. *Cancer Genomics and Proteomics* **2006,** 3, (5), 279.

4. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.;

Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; de Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y. J., Initial sequencing and analysis of the human genome. *Nature* **2001,** 409, (6822), 860-921.

5. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W., The sequence of the human genome. *Science* **2001,** 291, (5507), 1304-51.

6. Guerrera, I. C.; Kleiner, O., Application of mass spectrometry in proteomics. *Biosci Rep* **2005,** 25, (1-2), 71-93.

7. Caprioli, R. M., Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Res* **2005,** 65, (23), 10642-5.

8. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **2005,** 44, (45), 7342-72.

9. Hanash, S., HUPO initiatives relevant to clinical proteomics. *Mol Cell Proteomics* **2004,** 3, (4), 298-301.

10. Omenn, G. S., International collaboration in clinical chemistry and laboratory medicine: the Human Proteome Organization (HUPO) Plasma Proteome Project. *Clin Chem Lab Med* **2004,** 42, (1), 1-2.

11. Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C. A.; Amini, B.; Andersen, E.; Andersson, A. C.; Angelidou, P.; Asplund, A.; Asplund, C.; Berglund, L.; Bergstrom, K.; Brumer, H.; Cerjan, D.; Ekstrom, M.; Elobeid, A.; Eriksson, C.; Fagerberg, L.; Falk, R.; Fall, J.; Forsberg, M.; Bjorklund, M. G.; Gumbel, K.; Halimi, A.; Hallin, I.; Hamsten, C.; Hansson, M.; Hedhammar, M.; Hercules, G.; Kampf, C.; Larsson, K.; Lindskog, M.; Lodewyckx, W.; Lund, J.; Lundeberg, J.; Magnusson, K.; Malm, E.; Nilsson, P.; Odling, J.; Oksvold, P.; Olsson, I.; Oster, E.; Ottosson, J.; Paavilainen, L.; Persson, A.; Rimini, R.; Rockberg, J.; Runeson, M.; Sivertsson, A.; Skollermo, A.; Steen, J.; Stenvall, M.; Sterky, F.; Stromberg, S.; Sundberg, M.; Tegel, H.; Tourle, S.; Wahlund, E.; Walden, A.; Wan, J.; Wernerus, H.; Westberg, J.; Wester, K.; Wrethagen, U.; Xu, L. L.; Hober, S.;

Ponten, F., A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **2005,** 4, (12), 1920-32.

12. Roboz, J., Mass spectrometry in diagnostic oncoproteomics. *Cancer Invest* **2005,** 23, (5), 465-78.

13. Vidal, B. C.; Bonventre, J. V.; S, I. H. H., Towards the application of proteomics in renal disease diagnosis. *Clin Sci (Lond)* **2005,** 109, (5), 421-30.

14. Charrier, J. P.; Tournel, C.; Michel, S.; Comby, S.; Jolivet-Reynaud, C.; Passagot, J.; Dalbon, P.; Chautard, D.; Jolivet, M., Differential diagnosis of prostate cancer and benign prostate hyperplasia using two-dimensional electrophoresis. *Electrophoresis* **2001,** 22, (9), 1861-6.

15. Johnson, G.; Brane, D.; Block, W.; van Kammen, D. P.; Gurklis, J.; Peters, J. L.; Wyatt, R. J.; Kirch, D. G.; Ghanbari, H. A.; Merril, C. R., Cerebrospinal fluid protein variations in common to Alzheimer's disease and schizophrenia. *Appl Theor Electrophor* **1992,** 3, (2), 47-53.

16. Saunders, F. K.; Winfield, D. A.; Goepel, J. R.; Hancock, B. W.; Sharrard, R. M.; Goyns, M. H., 2D-gel analysis of protein synthesis profiles of different stages of chronic lymphocytic leukaemia. *Leuk Lymphoma* **1994,** 14, (3-4), 319-22.

17. Beer, H. L.; Jenkins, R. E.; Gutowska-Owsiak, D.; Pazmany, L.; Birchall, M. A.; Kitteringham, N. R., 2-Dimensional gel electrophoresis and MALDI-MS of cystic cervical metastasis from head and neck squamous cell carcinoma. *Clin Otolaryngol* **2006,** 31, (3), 246.

18. Lilley, K. S.; Friedman, D. B., All about DIGE: quantification technology for differential-display 2D-gel proteomics. *Expert Rev Proteomics* **2004,** 1, (4), 401-9.

19. Unlu, M.; Morgan, M. E.; Minden, J. S., Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **1997,** 18, (11), 2071-7.

20. Alban, A.; David, S. O.; Bjorkesten, L.; Andersson, C.; Sloge, E.; Lewis, S.; Currie, I., A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **2003,** 3, (1), 36-44.

21. Gharbi, S.; Gaffney, P.; Yang, A.; Zvelebil, M. J.; Cramer, R.; Waterfield, M. D.; Timms, J. F., Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Proteomics* **2002,** 1, (2), 91-8.

22. Zhou, G.; Li, H.; DeCamp, D.; Chen, S.; Shu, H.; Gong, Y.; Flaig, M.; Gillespie, J. W.; Hu, N.; Taylor, P. R.; Emmert-Buck, M. R.; Liotta, L. A.; Petricoin, E. F., 3rd; Zhao, Y., 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol Cell Proteomics* **2002,** 1, (2), 117-24.

23. Friedman, D. B.; Hill, S.; Keller, J. W.; Merchant, N. B.; Levy, S. E.; Coffey, R. J.; Caprioli, R. M., Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* **2004,** 4, (3), 793-811.

24. Link, A. J., Multidimensional peptide separations in proteomics. *Trends Biotechnol* **2002,** 20, (12 Suppl), S8-13.

25. Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.; Emili, A., Human Tissue Profiling with Multidimensional Protein Identification Technology. *J Proteome Res* **2005,** 4, (5), 1757-1767.

26. Florens, L.; Washburn, M. P., Proteomic analysis by multidimensional protein identification technology. *Methods Mol Biol* **2006,** 328, 159-75.

27. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999,** 17, (7), 676-82.

28. Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A., 3rd, Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal Biochem* **1998,** 258, (2), 349-61.

29. Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M., Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **2002,** 20, (3), 301-5.

30. Cargile, B. J.; Bundy, J. L.; Freeman, T. W.; Stephenson, J. L., Jr., Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. *J Proteome Res* **2004,** 3, (1), 112-9.

31. Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr., Potential for false positive identifications from large databases through tandem mass spectrometry. *J Proteome Res* **2004,** 3, (5), 1082-5.

32. Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Stephenson, J. L., Jr.; Bundy, J. L., Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. *J Biomol Tech* **2005,** 16, (3), 181-9.

33. Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr., A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* **2005,** 5, (1), 24-34.

34. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.

35. Zieske, L. R., A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* **2006,** 57, (7), 1501-8.

36. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.

37. Heller, M.; Mattou, H.; Menzel, C.; Yao, X., Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J Am Soc Mass Spectrom* **2003,** 14, (7), 704-18.

38. Cho, S. H.; Goodlett, D.; Franzblau, S., ICAT-based comparative proteomic analysis of non-replicating persistent Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **2005**.

39. Stewart, J. J.; White, J. T.; Yan, X.; Collins, S.; Drescher, C. W.; Urban, N. D.; Hood, L.; Lin, B., Proteins associated with cisplatin resistance in ovarian cancer cells

identified by quantitative proteomics technology and integrated with mRNA expression levels. *Mol Cell Proteomics* **2005**.

40. Keshamouni, V. G.; Michailidis, G.; Grasso, C. S.; Anthwal, S.; Strahler, J. R.; Walker, A.; Arenberg, D. A.; Reddy, R. C.; Akulapalli, S.; Thannickal, V. J.; Standiford, T. J.; Andrews, P. C.; Omenn, G. S., Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J Proteome Res* **2006,** 5, (5), 1143-54.

41. DeSouza, L.; Diehl, G.; Rodrigues, M. J.; Guo, J.; Romaschin, A. D.; Colgan, T. J.; Siu, K. W., Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J Proteome Res* **2005,** 4, (2), 377-86.

42. Everley, P. A.; Krijgsveld, J.; Zetter, B. R.; Gygi, S. P., Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol Cell Proteomics* **2004,** 3, (7), 729-35.

43. Fang, R.; Elias, D. A.; Monroe, M. E.; Shen, Y.; McIntosh, M.; Wang, P.; Goddard, C. D.; Callister, S. J.; Moore, R. J.; Gorby, Y. A.; Adkins, J. N.; Fredrickson, J. K.; Lipton, M. S.; Smith, R. D., Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol Cell Proteomics* **2006,** 5, (4), 714-25.

44. Yates, J. R., 3rd, Mass spectrometry and the age of the proteome. *J Mass Spectrom* **1998,** 33, (1), 1-19.

45. Zenobi, R.; Knochenmuss, R., Ion Formation in MALDI Mass Spectrometry. *Mass Spectrometry Reviews* **1998,** 17, 337-366.

46. Karas, M.; Gluckmann, M.; Schafer, J., Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *Journal of Mass Spectrometry* **2000,** 35, (1), 1-12.

47. Wiley, W. C.; McLaren, I. H., Time-of-Flight Mass Spectrometer with Improved Resolution. *The Review of Scientific Instruments* **1955,** 26, (12), 1150-1157.

48. Boccaccio, P.; Vannucci, L.; Ricci, R. A.; Massa, I.; Vannini, G.; Sarti, A., On the Detection Efficiency of a Microchannel Plate Detector in Time-of-Flight Measurements.

Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment **1986,** 243, (2-3), 599-600.

49. Wiza, J. L., Microchannel Plate Detectors. *Nuclear Instruments & Methods* **1979,** 162, (1-3), 587-601.

50. Cotter, R. J., The new time-of-flight mass spectrometry. *Analytical Chemistry* **1999,** 71, (13), 445A-451A.

51. Standing, K. G., Timing the flight of biomolecules: a personal perspective. *International Journal of Mass Spectrometry* **2000,** 200, (1-3), 597-610.

52. Mamyrin, B. A., Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry* **2001,** 206, (3), 251-266.

53. 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.

54. Spengler, B.; Hubert, M., Scanning microprobe matrix-assisted laser desorption ionization (SMALDI) mass spectrometry: instrumentation for sub-micrometer resolved LDI and MALDI surface analysis. *J Am Soc Mass Spectrom* **2002,** 13, (6), 735-48.

55. McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R., Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal Chem* **2005,** 77, (19), 6118-24.

56. Stoeckli, M.; Staab, D.; Staufenbiel, M.; Wiederhold, K. H.; Signor, L., Molecular imaging of amyloid beta peptides in mouse brain sections using mass spectrometry. *Anal Biochem* **2002,** 311, (1), 33-9.

57. 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.

58. Touboul, D.; Piednoel, H.; Voisin, V.; De La Porte, S.; Brunelle, A.; Halgand, F.; Laprevote, O., Changes of phospholipid composition within the dystrophic muscle by matrix-assisted laser desorption/ionization mass spectrometry and mass spectrometry imaging. *Eur J Mass Spectrom (Chichester, Eng)* **2004,** 10, (5), 657-64.

59. Reyzer, M. L.; Caldwell, R. L.; Dugger, T. C.; Forbes, J. T.; Ritter, C. A.; Guix, M.; Arteaga, C. L.; Caprioli, R. M., Early changes in protein expression detected by mass spectrometry predict tumor response to molecular therapeutics. *Cancer Res* **2004,** 64, (24), 9093-100.

60. 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.

61. Chaurand, P.; Stoeckli, M.; Caprioli, R. M., Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. *Anal Chem* **1999,** 71, (23), 5263-70.

62. Jimenez, C. R.; Burlingame, A. L., Ultramicroanalysis of peptide profiles in biological samples using MALDI mass spectrometry. *Exp Nephrol* **1998,** 6, (5), 421-8.

63. 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.

64. Redeker, V.; Toullec, J. Y.; Vinh, J.; Rossier, J.; Soyez, D., Combination of peptide profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and immunodetection on single glands or cells. *Anal Chem* **1998,** 70, (9), 1805-11.

65. 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.

66. 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. *Nat Med* **2001,** 7, (4), 493-6.

67. 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), 7674-81.

68. Xu, B. J.; Shyr, Y.; Liang, X.; Ma, L. J.; Donnert, E. M.; Roberts, J. D.; Zhang, X.; Kon, V.; Brown, N. J.; Caprioli, R. M.; Fogo, A. B., Proteomic patterns and prediction of glomerulosclerosis and its mechanisms. *J Am Soc Nephrol* **2005,** 16, (10), 2967-75.

69. Pierson, J.; Norris, J. L.; Aerni, H. R.; Svenningsson, P.; Caprioli, R. M.; Andren, P. E., Molecular profiling of experimental Parkinson's disease: direct analysis of peptides and proteins on brain tissue sections by MALDI mass spectrometry. *J Proteome Res* **2004,** 3, (2), 289-95.

70. 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.

71. Yanagisawa, K.; Xu, B. J.; Carbone, D. P.; Caprioli, R. M., Molecular fingerprinting in human lung cancer. *Clin Lung Cancer* **2003,** 5, (2), 113-8.

72. Xie, L.; Xu, B. J.; Gorska, A. E.; Shyr, Y.; Schwartz, S. A.; Cheng, N.; Levy, S.; Bierie, B.; Caprioli, R. M.; Moses, H. L., Genomic and proteomic analysis of mammary tumors arising in transgenic mice. *J Proteome Res* **2005,** 4, (6), 2088-98.

73. Belu, A. M.; Graham, D. J.; Castner, D. G., Time-of-flight secondary ion mass spectrometry: techniques and applications for the characterization of biomaterial surfaces. *Biomaterials* **2003,** 24, (21), 3635-53.

74. 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.

75. Touboul, D.; Brunelle, A.; Halgand, F.; De La Porte, S.; Laprevote, O., Lipid imaging by gold cluster time-of-flight secondary ion mass spectrometry: application to Duchenne muscular dystrophy. *J Lipid Res* **2005,** 46, (7), 1388-95.

76. Jemal, A.; Murray, T.; Ward, E.; Samuels, A.; Tiwari, R.; Ghafoor, A.; Feuer, E.; Thun, M., Cancer Statistics, 2005. *CA Cancer J Clin* **2005,** 55, 10-30.

77. Kosari, F.; Parker, A. S.; Kube, D. M.; Lohse, C. M.; Leibovich, B. C.; Blute, M. L.; Cheville, J. C.; Vasmatzis, G., Clear cell renal cell carcinoma: gene expression

analyses identify a potential signature for tumor aggressiveness. *Clin Cancer Res* **2005,** 11, (14), 5128-39.

78. Murphy, W.; Grignon, D. J.; Perlman, E., *Tumors of the kidney, bladder, and related urinary structures*. American Registry of Pathology: Washington, D.C., 2004.

79. Linehan, W. M.; Vasselli, J.; Srinivasan, R.; Walther, M. M.; Merino, M.; Choyke, P.; Vocke, C.; Schmidt, L.; Isaacs, J. S.; Glenn, G.; Toro, J.; Zbar, B.; Bottaro, D.; Neckers, L., Genetic basis of cancer of the kidney: disease-specific approaches to therapy. *Clin Cancer Res* **2004,** 10, (18 Pt 2), 6282S-9S.

80. Linehan, W. M.; Lerman, M. I.; Zbar, B., Identification of the von Hippel-Lindau (VHL) gene. Its role in renal cancer. *Jama* **1995,** 273, (7), 564-70.

81. Poston, C. D.; Jaffe, G. S.; Lubensky, I. A.; Solomon, D.; Zbar, B.; Linehan, W. M.; Walther, M. M., Characterization of the renal pathology of a familial form of renal cell carcinoma associated with von Hippel-Lindau disease: clinical and molecular genetic implications. *J Urol* **1995,** 153, (1), 22-6.

82. Walther, M. M.; Lubensky, I. A.; Venzon, D.; Zbar, B.; Linehan, W. M., Prevalence of microscopic lesions in grossly normal renal parenchyma from patients with von Hippel-Lindau disease, sporadic renal cell carcinoma and no renal disease: clinical implications. *J Urol* **1995,** 154, (6), 2010-4; discussion 2014-5.

83. Anglard, P.; Tory, K.; Brauch, H.; Weiss, G. H.; Latif, F.; Merino, M. J.; Lerman, M. I.; Zbar, B.; Linehan, W. M., Molecular analysis of genetic changes in the origin and development of renal cell carcinoma. *Cancer Res* **1991,** 51, (4), 1071-7.

84. Kovacs, G.; Szucs, S.; De Riese, W.; Baumgartel, H., Specific chromosome aberration in human renal cell carcinoma. *Int J Cancer* **1987,** 40, (2), 171-8.

85. Yoshida, M. A.; Ohyashiki, K.; Ochi, H.; Gibas, Z.; Pontes, J. E.; Prout, G. R., Jr.; Huben, R.; Sandberg, A. A., Cytogenetic studies of tumor tissue from patients with nonfamilial renal cell carcinoma. *Cancer Res* **1986,** 46, (4 Pt 2), 2139-47.

86. Diaz, J. I.; Mora, L. B.; Hakam, A., The Mainz Classification of Renal Cell Tumors. *Cancer Control* **1999,** 6, (6), 571-579.

87. Levin, H. S.; Myles, J. L., *Renal cell carcinoma: molecular biology, immunology, and clinical management*. Humana Press: Totowa, N.J., 2000; Vol. 4.

88. Dechet, C. B.; Zincke, H.; Sebo, T. J.; King, B. F.; LeRoy, A. J.; Farrow, G. M.; Blute, M. L., Prospective analysis of computerized tomography and needle biopsy with permanent sectioning to determine the nature of solid renal masses in adults. *J Urol* **2003,** 169, (1), 71-4.

89. Sengupta, S.; Zincke, H., Lessons learned in the surgical management of renal cell carcinoma. *Urology* **2005,** 66, (5 Suppl), 36-42.

90. Fuhrman, S. A.; Lasky, L. C.; Limas, C., Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol* **1982,** 6, (7), 655-63.

91. Mejean, A.; Oudard, S.; Thiounn, N., Prognostic factors of renal cell carcinoma. *J Urol* **2003,** 169, (3), 821-7.

92. Altman, D. G.; Lausen, B.; Sauerbrei, W.; Schumacher, M., Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* **1994,** 86, (11), 829-35.

93. Fielding, L. P.; Henson, D. E., Multiple prognostic factors and outcome analysis in patients with cancer. Communication from the American Joint Committee on Cancer. *Cancer* **1993,** 71, (7), 2426-9.

94. Gelb, A. B., Renal cell carcinoma: current prognostic factors. Union Internationale Contre le Cancer (UICC) and the American Joint Committee on Cancer (AJCC). *Cancer* **1997,** 80, (5), 981-6.

95. Schuetz, A. N.; Yin-Goen, Q.; Amin, M. B.; Moreno, C. S.; Cohen, C.; Hornsby, C. D.; Yang, W. L.; Petros, J. A.; Issa, M. M.; Pattaras, J. G.; Ogan, K.; Marshall, F. F.; Young, A. N., Molecular classification of renal tumors by gene expression profiling. *J Mol Diagn* **2005,** 7, (2), 206-18.

96. Yao, M.; Tabuchi, H.; Nagashima, Y.; Baba, M.; Nakaigawa, N.; Ishiguro, H.; Hamada, K.; Inayama, Y.; Kishida, T.; Hattori, K.; Yamada-Okabe, H.; Kubota, Y., Gene expression analysis of renal carcinoma: adipose differentiation-related protein as a potential diagnostic and prognostic biomarker for clear-cell renal carcinoma. *J Pathol* **2005,** 205, (3), 377-87.

97. Pizem, J.; Cor, A.; Gale, N., Survivin expression is a negative prognostic marker in laryngeal squamous cell carcinoma and is associated with p53 accumulation. *Histopathology* **2004,** 45, (2), 180-6.

98. Shariat, S. F.; Lotan, Y.; Saboorian, H.; Khoddami, S. M.; Roehrborn, C. G.; Slawin, K. M.; Ashfaq, R., Survivin expression is associated with features of biologically aggressive prostate carcinoma. *Cancer* **2004,** 100, (4), 751-7.

99. Kren, L.; Brazdil, J.; Hermanova, M.; Goncharuk, V. N.; Kallakury, B. V.; Kaur, P.; Ross, J. S., Prognostic significance of anti-apoptosis proteins survivin and bcl-2 in non-small cell lung carcinomas: a clinicopathologic study of 102 cases. *Appl Immunohistochem Mol Morphol* **2004,** 12, (1), 44-9.

100. Miyachi, K.; Sasaki, K.; Onodera, S.; Taguchi, T.; Nagamachi, M.; Kaneko, H.; Sunagawa, M., Correlation between survivin mRNA expression and lymph node metastasis in gastric cancer. *Gastric Cancer* **2003,** 6, (4), 217-24.

101. Sarto, C.; Marocchi, A.; Sanchez, J. C.; Giannone, D.; Frutiger, S.; Golaz, O.; Wilkins, M. R.; Doro, G.; Cappellano, F.; Hughes, G.; Hochstrasser, D. F.; Mocarelli, P., Renal cell carcinoma and normal kidney protein expression. *Electrophoresis* **1997,** 18, (3-4), 599-604.

102. Seliger, B.; Lichtenfels, R.; Atkins, D.; Bukur, J.; Halder, T.; Kersten, M.; Harder, A.; Ackermann, A.; Malenica, B.; Brenner, W.; Zobawa, M.; Lottspeich, F., Identification of fatty acid binding proteins as markers associated with the initiation and/or progression of renal cell carcinoma. *Proteomics* **2005,** 5, (10), 2631-40.

103. Hwa, J. S.; Park, H. J.; Jung, J. H.; Kam, S. C.; Park, H. C.; Kim, C. W.; Kang, K. R.; Hyun, J. S.; Chung, K. H., Identification of proteins differentially expressed in the conventional renal cell carcinoma by proteomic analysis. *J Korean Med Sci* **2005,** 20, (3), 450-5.

104. Seliger, B.; Lichtenfels, R.; Kellner, R., Detection of renal cell carcinoma-associated markers via proteome- and other 'ome'-based analyses. *Brief Funct Genomic Proteomic* **2003,** 2, (3), 194-212.

105. Sarto, C.; Deon, C.; Doro, G.; Hochstrasser, D. F.; Mocarelli, P.; Sanchez, J. C., Contribution of proteomics to the molecular analysis of renal cell carcinoma with an emphasis on manganese superoxide dismutase. *Proteomics* **2001,** 1, (10), 1288-94.

106. Looser, K. G.; Shah, J. P.; Strong, E. W., The significance of "positive" margins in surgically resected epidermoid carcinomas. *Head Neck Surg* **1978,** 1, (2), 107-11.

107. Nathan, C.-A. O.; Amirghahri, N.; Rice, C.; Abreo, F.; Shi, R.; Stucker, S., Molecular Analysis of Surgical Margins in Head and Neck Squamous Cell Carcinoma Patients. *The Laryngoscope* **2002,** 112, 2129-2140.

108. Masasyesva, B. G.; Tong, B. C.; Brock, M. V.; Pilkington, T.; Goldenberg, D.; Sidransky, D.; Harden, S.; Westra, W. H.; Califano, J., Molecular margin analysis predicts local recurrence after sublobar resection of lung cancer. *Int. J. Cancer* **2005,** 113, (1022-1025).

109. Balch, G. C.; Mithani, S. K.; Simpson, J. F.; Kelley, M. C., Accuracy of intraoperative gross examination of surgical margin status in women undergoing partial mastectomy for breast malignancy. *Am Surg* **2005,** 71, (1), 22-7; discussion 27-8.

110. Choong, P. F., Surgical margins for soft tissue sarcoma: size does matter. *ANZ J Surg* **2006,** 76, (3), 97.

111. Gronchi, A.; Casali, P. G.; Mariani, L.; Miceli, R.; Fiore, M.; Lo Vullo, S.; Bertulli, R.; Collini, P.; Lozza, L.; Olmi, P.; Rosai, J., Status of surgical margins and prognosis in adult soft tissue sarcomas of the extremities: a series of patients treated at a single institution. *J Clin Oncol* **2005,** 23, (1), 96-104.

112. van Houten, V. M.; Leemans, C. R.; Kummer, J. A.; Dijkstra, J.; Kuik, D. J.; van den Brekel, M. W.; Snow, G. B.; Brakenhoff, R. H., Molecular diagnosis of surgical margins and local recurrence in head and neck cancer patients: a prospective study. *Clin Cancer Res* **2004,** 10, (11), 3614-20.

113. Toms, S. A.; Lin, W. C.; Weil, R. J.; Johnson, M. D.; Jansen, E. D.; Mahadevan-Jansen, A., Intraoperative optical spectroscopy identifies infiltrating glioma margins with high sensitivity. *Neurosurgery* **2005,** 57, (4 Suppl), 382-91; discussion 382-91.

114. Novick, A. C., Partial nephrectomy for renal cell carcinoma. *Urol Clin North Am* **1987,** 14, (2), 419-33.

115. Castilla, E. A.; Liou, L. S.; Abrahams, N. A.; Fergany, A.; Rybicki, L. A.; Myles, J.; Novick, A. C., Prognostic importance of resection margin width after nephron-sparing surgery for renal cell carcinoma. *Urology* **2002,** 60, (6), 993-7.

116. Lerner, S. E.; Hawkins, C. A.; Blute, M. L.; Grabner, A.; Wollan, P. C.; Eickholt, J. T.; Zincke, H., Disease outcome in patients with low stage renal cell carcinoma treated with nephron sparing or radical surgery. *J Urol* **1996,** 155, (6), 1868-73.

117. Sutherland, S. E.; Resnick, M. I.; Maclennan, G. T.; Goldman, H. B., Does the size of the surgical margin in partial nephrectomy for renal cell cancer really matter? *J Urol* **2002,** 167, (1), 61-4.

118. Nguyen, T. T.; Parkinson, J. P.; Kuehn, D. M.; Winfield, H. N., Technique for ensuring negative surgical margins during laparoscopic partial nephrectomy. *J Endourol* **2005,** 19, (3), 410-5.

119. Karlsson, J. O.; Toner, M., Long-term storage of tissues by cryopreservation: critical issues. *Biomaterials* **1996,** 17, (3), 243-56.

120. Mazur, P., Freezing of living cells: mechanisms and implications. *Am J Physiol* **1984,** 247, (3 Pt 1), C125-42.

121. Grizzle, W. E.; Aamodt, R.; Clausen, K.; LiVolsi, V.; Pretlow, T. G.; Qualman, S., Providing human tissues for research: how to establish a program. *Arch Pathol Lab Med* **1998,** 122, (12), 1065-76.

122. 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.

123. 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.

124. McCombie, G.; Knochenmuss, R., Enhanced MALDI ionization efficiency at the metal-matrix interface: practical and mechanistic consequences of sample thickness and preparation method. *J Am Soc Mass Spectrom* **2006,** 17, (5), 737-45.

125. 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.

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

127. Beavis, R. C.; Chait, B. T., Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun Mass Spectrom* **1989,** 3, (12), 432-5.

128. Beavis, R. C.; Chait, B. T., Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc Natl Acad Sci U S A* **1990,** 87, (17), 6873-7.

129. Srinivasan, M.; Sedmak, D.; Jewell, S., Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* **2002,** 161, (6), 1961-71.

130. Chaurand, P.; Cornett, D. S.; Caprioli, R. M., Molecular imaging of thin mammalian tissue sections by mass spectrometry. *Curr Opin Biotechnol* **2006**.

131. 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.

132. Nakanishi, T.; Ohtsu, I.; Furuta, M.; Ando, E.; Nishimura, O., Direct MS/MS analysis of proteins blotted on membranes by a matrix-assisted laser desorption/ionization-quadrupole ion trap-time-of-flight tandem mass spectrometer. *J Proteome Res* **2005,** 4, (3), 743-7.

133. Groseclose, M. R.; Andersson, M.; Hardesty, W. H.; Caprioli, R. M., Identification of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry. *Journal of Mass Spectrometry* **2006,** In Press.

134. 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.

135. 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.

136. Hu, J.; Coombes, K. R.; Morris, J. S.; Baggerly, K. A., The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. *Brief Funct Genomic Proteomic* **2005,** 3, (4), 322-31.

137. Holle, A.; Haase, A.; Kayser, M.; Hohndorf, J., Optimizing UV laser focus profiles for improved MALDI performance. *Journal of Mass Spectrometry* **2006,** 41, (6), 705-716.

138. Norris, J. L.; Cornett, D. S.; Mobley, J. A.; Andersson, M.; Seeley, E. H.; Chaurand, P.; Caprioli, R. M., Processing MALDI mass spectra to aid biomarker discovery and improve mass spectral image quality. *International Journal of Mass Spectrometry* **2006,** Submitted.

139. Wagner, M.; Naik, D.; Pothen, A., Protocols for disease classification from mass spectrometry data. *Proteomics* **2003,** 3, 1692-1698.

140. Baggerly, K. A.; Morris, J. S.; Wang, J.; Gold, D.; Xiao, L.; Coombes, K. R., A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. *Proteomics* **2003,** 3, 1667-1672.

141. Coombes, K. R.; Koomen, J. M.; Baggerly, K. A.; Morris, J. S.; Kobayashi, R., Understanding the characteristics of mass spectrometry data through the use of simulation. *Cancer Informatics* **2005,** 1, (1), 41-52.

142. Hilario, M.; Kalousis, A.; Pellegrini, C.; Muller, M., Processing and classification of protein mass spectra. *Mass Spectrometry Reviews* **2006,** 25, 409-449.

143. 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), 5116-5121.

144. Naiman, D. Q., Random Data Set Generation to Support Microarray Analysis. *Methods in Enzymology* **2006,** 411.

145. Hedenfalk, I.; Duggan, D.; Chen, Y.; Radmacher, M.; Bittner, M.; Simon, R.; Meltzer, P.; Gusterson, B.; Esteller, M.; Kallioniemi, O. P.; Wilfond, B.; Borg, A.; Trent, J.; Raffeld, M.; Yakhini, Z.; Ben-Dor, A.; Dougherty, E.; Kononen, J.; Bubendorf, L.; Fehrle, W.; Pittaluga, S.; Gruvberger, S.; Loman, N.; Johannsson, O.; Olsson, H.; Sauter, G., Gene-expression profiles in hereditary breast cancer. *N Engl J Med* **2001,** 344, (8), 539-48.

146. Storey, J. D.; Tibshirani, R., Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **2003,** 100, (16), 9440-5.

147. Harrell, F. E., Jr., *Regression Modeling Strategies*. Springer: New York, NY, 2001.

148. Ihaka, R.; Gentleman, R., R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **1996,** 5, (3), 299-314.

149. Slaughter, D. P.; Southwick, H. W.; Smejkal, W., Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* **1953,** 6, (5), 963-8.

150. Copper, M. P.; Braakhuis, B. J.; de Vries, N.; van Dongen, G. A.; Nauta, J. J.; Snow, G. B., A panel of biomarkers of carcinogenesis of the upper aerodigestive tract as potential intermediate endpoints in chemoprevention trials. *Cancer* **1993,** 71, (3), 825-30.

151. Stern, R. S.; Bolshakov, S.; Nataraj, A. J.; Ananthaswamy, H. N., p53 mutation in nonmelanoma skin cancers occurring in psoralen ultraviolet a-treated patients: evidence for heterogeneity and field cancerization. *J Invest Dermatol* **2002,** 119, (2), 522-6.

152. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **2002,** 74, (20), 5383-92.

153. Teratani, T.; Watanabe, T.; Kuwahara, F.; Kumagai, H.; Kobayashi, S.; Aoki, U.; Ishikawa, A.; Arai, K.; Nozawa, R., Induced transcriptional expression of calcium-binding protein S100A1 and S100A10 genes in human renal cell carcinoma. *Cancer Lett* **2002,** 175, (1), 71-7.

154. Domoto, T.; Miyama, Y.; Suzuki, H.; Teratani, T.; Arai, K.; Sugiyama, T.; Takayama, T.; Mugiya, S.; Ozono, S.; Nozawa, R., Evaluation of S100A10, annexin II and B-FABP expression as markers for renal cell carcinoma. *Cancer Sci* **2006**.

155. Madej, A.; Puzianowska-Kuznicka, M.; Tanski, Z.; Nauman, J.; Nauman, A., Vitamin D receptor binding to DNA is altered without the change in its expression in human renal clear cell cancer. *Nephron Exp Nephrol* **2003,** 93, (4), e150-7.

156. Luu, H. H.; Zhou, L.; Haydon, R. C.; Deyrup, A. T.; Montag, A. G.; Huo, D.; Heck, R.; Heizmann, C. W.; Peabody, T. D.; Simon, M. A.; He, T. C., Increased expression of S100A6 is associated with decreased metastasis and inhibition of cell migration and anchorage independent growth in human osteosarcoma. *Cancer Lett* **2005,** 229, (1), 135-48.

157. Rehman, I.; Cross, S. S.; Azzouzi, A. R.; Catto, J. W.; Deloulme, J. C.; Larre, S.; Champigneuille, J.; Fromont, G.; Cussenot, O.; Hamdy, F. C., S100A6 (Calcyclin) is a prostate basal cell marker absent in prostate cancer and its precursors. *Br J Cancer* **2004,** 91, (4), 739-44.

158. Hall, A. K., Amplification-independent overexpression of thymosin beta-10 mRNA in human renal cell carcinoma. *Ren Fail* **1994,** 16, (2), 243-54.

159. Santelli, G.; Califano, D.; Chiappetta, G.; Vento, M. T.; Bartoli, P. C.; Zullo, F.; Trapasso, F.; Viglietto, G.; Fusco, A., Thymosin beta-10 gene overexpression is a general event in human carcinogenesis. *Am J Pathol* **1999,** 155, (3), 799-804.

160. Ren, Y.; Tsui, H. T.; Poon, R. T.; Ng, I. O.; Li, Z.; Chen, Y.; Jiang, G.; Lau, C.; Yu, W. C.; Bacher, M.; Fan, S. T., Macrophage migration inhibitory factor: roles in regulating tumor cell migration and expression of angiogenic factors in hepatocellular carcinoma. *Int J Cancer* **2003,** 107, (1), 22-9.

161. Mitchell, R. A.; Bucala, R., Tumor growth-promoting properties of macrophage migration inhibitory factor (MIF). *Semin Cancer Biol* **2000,** 10, (5), 359-66.

162. White, E. S.; Strom, S. R.; Wys, N. L.; Arenberg, D. A., Non-small cell lung cancer cells induce monocytes to increase expression of angiogenic activity. *J Immunol* **2001,** 166, (12), 7549-55.

163. Meyer-Siegler, K. L.; Bellino, M. A.; Tannenbaum, M., Macrophage migration inhibitory factor evaluation compared with prostate specific antigen as a biomarker in patients with prostate carcinoma. *Cancer* **2002,** 94, (5), 1449-56.

164. Simonnet, H.; Alazard, N.; Pfeiffer, K.; Gallou, C.; Beroud, C.; Demont, J.; Bouvier, R.; Schagger, H.; Godinot, C., Low mitochondrial respiratory chain content correlates with tumor aggressiveness in renal cell carcinoma. *Carcinogenesis* **2002,** 23, (5), 759-68.

165. 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), 1057-68.

166. Johnson, M. D.; Floyd, J. L.; Caprioli, R. M., Proteomics in diagnostic neuropathology. *J Neuropathol Exp Neurol* **2006,** 65, (9), 837-45.

167. Predki, P. F.; Mattoon, D.; Bangham, R.; Schweitzer, B.; Michaud, G., Protein microarrays: a new tool for profiling antibody cross-reactivity. *Hum Antibodies* **2005,** 14, (1-2), 7-15.

168. Hood, B. L.; Conrads, T. P.; Veenstra, T. D., Mass spectrometric analysis of formalin-fixed paraffin-embedded tissue: unlocking the proteome within. *Proteomics* **2006,** 6, (14), 4106-14.

169. 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), 1741-53.

170. Chung, J. Y.; Braunschweig, T.; Hewitt, S. M., Optimization of recovery of RNA from formalin-fixed, paraffin-embedded tissue. *Diagn Mol Pathol* **2006,** 15, (4), 229-36.

171. Palmer-Toy, D. E.; Krastins, B.; Sarracino, D. A.; Nadol, J. B., Jr.; Merchant, S. N., Efficient method for the proteomic analysis of fixed and embedded tissues. *J Proteome Res* **2005,** 4, (6), 2404-11.

172. Stanta, G.; Mucelli, S. P.; Petrera, F.; Bonin, S.; Bussolati, G., A novel fixative improves opportunities of nucleic acids and proteomic analysis in human archive's tissues. *Diagn Mol Pathol* **2006,** 15, (2), 115-23.

173. Hervouet, E.; Demont, J.; Pecina, P.; Vojtiskova, A.; Houstek, J.; Simonnet, H.; Godinot, C., A new role for the von Hippel-Lindau tumor suppressor protein: stimulation of mitochondrial oxidative phosphorylation complex biogenesis. *Carcinogenesis* **2005,** 26, (3), 531-9.

174. Lodish, H.; Berk, A.; Zipursky, L. S.; Matsudaira, P.; Baltimore, D.; Darnell, J., *Molecular Cell Biology*. Fourth ed.; W. H. Freeman and Company: New York, NY, 2000.

175. Guzy, R. D.; Schumacker, P. T., Oxygen sensing by mitochondria at complex III: the paradox of increased reactive oxygen species during hypoxia. *Exp Physiol* **2006,** 91, (5), 807-19.

176. Suto, D.; Sato, K.; Ohba, Y.; Yoshimura, T.; Fujii, J., Suppression of the pro-apoptotic function of cytochrome c by singlet oxygen via a haem redox state-independent mechanism. *Biochem J* **2005,** 392, (Pt 2), 399-406.

177. Lenka, N.; Vijayasarathy, C.; Mullick, J.; Avadhani, N. G., Structural organization and transcription regulation of nuclear genes encoding the mammalian cytochrome c oxidase complex. *Prog Nucleic Acid Res Mol Biol* **1998,** 61, 309-44.

178. Carroll, J.; Fearnley, I. M.; Shannon, R. J.; Hirst, J.; Walker, J. E., Analysis of the subunit composition of complex I from bovine heart mitochondria. *Mol Cell Proteomics* **2003,** 2, (2), 117-26.

179. Carroll, J.; Shannon, R. J.; Fearnley, I. M.; Walker, J. E.; Hirst, J., Definition of the nuclear encoded protein composition of bovine heart mitochondrial complex I. Identification of two new subunits. *J Biol Chem* **2002,** 277, (52), 50311-7.

180. Hirst, J.; Carroll, J.; Fearnley, I. M.; Shannon, R. J.; Walker, J. E., The nuclear encoded subunits of complex I from bovine heart mitochondria. *Biochim Biophys Acta* **2003,** 1604, (3), 135-50.

181. Pocsfalvi, G.; Cuccurullo, M.; Schlosser, G.; Cacace, G.; Siciliano, R. A.; Mazzeo, M. F.; Scacco, S.; Cocco, T.; Gnoni, A.; Malorni, A.; Papa, S., Shotgun proteomics for the characterization of subunit composition of mitochondrial complex I. *Biochim Biophys Acta* **2006,** 1757, (9-10), 1438-50.

182. Warburg, O., The metabolism of tumors. *London: Constable Press* **1930**.

183. Meierhofer, D.; Mayr, J. A.; Foetschl, U.; Berger, A.; Fink, K.; Schmeller, N.; Hacker, G. W.; Hauser-Kronberger, C.; Kofler, B.; Sperl, W., Decrease of mitochondrial DNA content and energy metabolism in renal cell carcinoma. *Carcinogenesis* **2004,** 25, (6), 1005-10.

184. Kaelin, W. G., Jr., The von hippel-lindau tumor suppressor protein and clear cell renal carcinoma. *Clin Cancer Res* **2007,** 13, (2), 680s-4s.

185. Acker, T.; Fandrey, J.; Acker, H., The good, the bad and the ugly in oxygen-sensing: ROS, cytochromes and prolyl-hydroxylases. *Cardiovasc Res* **2006,** 71, (2), 195-207.

186. Giaccia, A. J.; Simon, M. C.; Johnson, R., The biology of hypoxia: the role of oxygen sensing in development, normal function, and disease. *Genes Dev* **2004,** 18, (18), 2183-94.

187. Haase, V. H., The VHL/HIF oxygen-sensing pathway and its relevance to kidney disease. *Kidney Int* **2006,** 69, (8), 1302-7.

188. Gatenby, R. A.; Gawlinski, E. T., The glycolytic phenotype in carcinogenesis and tumor invasion: insights through mathematical models. *Cancer Res* **2003,** 63, (14), 3847-54.

189. Raghunand, N.; Gatenby, R. A.; Gillies, R. J., Microenvironmental and cellular consequences of altered blood flow in tumours. *Br J Radiol* **2003,** 76 Spec No 1, S11-22.

190. Jiang, B. H.; Semenza, G. L.; Bauer, C.; Marti, H. H., Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of O2 tension. *Am J Physiol* **1996,** 271, (4 Pt 1), C1172-80.

191. Graziewicz, M. A.; Day, B. J.; Copeland, W. C., The mitochondrial DNA polymerase as a target of oxidative damage. *Nucleic Acids Res* **2002,** 30, (13), 2817-24.

192. Mekhail, K.; Gunaratnam, L.; Bonicalzi, M. E.; Lee, S., HIF activation by pH-dependent nucleolar sequestration of VHL. *Nat Cell Biol* **2004,** 6, (7), 642-7.

193. Pan, J.; Mestas, J.; Burdick, M. D.; Phillips, R. J.; Thomas, G. V.; Reckamp, K.; Belperio, J. A.; Strieter, R. M., Stromal derived factor-1 (SDF-1/CXCL12) and CXCR4 in renal cell carcinoma metastasis. *Mol Cancer* **2006,** 5, 56.

194. Staller, P.; Sulitkova, J.; Lisztwan, J.; Moch, H.; Oakeley, E. J.; Krek, W., Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. *Nature* **2003,** 425, (6955), 307-11.

195. Petrella, B. L.; Brinckerhoff, C. E., Tumor cell invasion of von Hippel Lindau renal cell carcinoma cells is mediated by membrane type-1 matrix metalloproteinase. *Mol Cancer* **2006,** 5, 66.

196. Crecelius, A. C.; Cornett, D. S.; Caprioli, R. M.; Williams, B.; Dawant, B. M.; Bodenheimer, B., Three-dimensional visualization of protein expression in mouse brain structures using imaging mass spectrometry. *J Am Soc Mass Spectrom* **2005,** 16, (7), 1093-9.