MOLECULAR ANALYSIS OF CANCER PROGRESSION WITH LABEL-FREE RAMAN SPECTROSCOPY

by

Santosh Kumar Paidi

A dissertation submitted to Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
April 2020

Abstract

Due to its ability to probe water-containing samples using visible and near-infrared frequencies with high chemical specificity, Raman spectroscopy is an attractive tool for label-free investigation of biological samples. While Raman spectroscopy has been leveraged for exploratory studies in clinical cancer diagnostics, only limited studies have used it to understand the molecular mechanisms driving key characteristics of cancer progression. In this thesis, we present three progressively complex applications of Raman spectroscopy that take advantage of its specificity and synergistic combination with plasmonic nanoparticles and multivariate data analysis for molecular study of cancer.

First, we used Au@SiO_2 shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS) to investigate the roles of microcalcification status and the composition of tumor microenvironment in breast tissue for identification of a range of breast pathologies. We developed a partial least squares-discriminant analysis-based classifier to correlate the spectra with their pathology to obtain high prediction accuracy. A parallel investigation of the genetic drivers of microcalcification formation in breast cancer cells revealed that stable silencing of the Osteopontin gene decreased the formation of hydroxyapatite in breast cancer cells and reduced their migration.

Next, we demonstrated the ability to detect premetastatic changes in the lungs of mice bearing breast tumors, in advance of tumor cell seeding, using Raman spectroscopy and multivariate data analysis. Our measurements showed reliable differences in the collagen and proteoglycan features of the premetastatic lungs which uniquely identify the metastatic potential of the primary tumor. Consistent with histological assessment, our results hint at a continuous premetastatic niche formation model dependent on the metastatic potential of primary tumor.

Finally, we exploited Raman mapping to elucidate radiation therapy-induced biomolecular changes in murine tumors and uncovered latent microenvironmental differences between treatment-resistant and -sensitive tumors. We used multivariate curve resolution-alternating least squares (MCR-ALS) and support vector machine (SVM) to quantify biomolecular differences in the tumor microenvironment and constructed classification models to predict therapy outcome and resistance. We found significant differences in lipid and collagen content between unirradiated and irradiated tumors.

Taken together, these studies pave the way for applications of Raman spectroscopy beyond clinical diagnostics such as metastatic risk assessment and treatment monitoring.

Advisor: Dr. Ishan Barman
Dissertation Readers: Dr. Tza-Huei (Jeff) Wang, Dr. Kristine Glunde, and
Dr. Yun Chen

Acknowledgments

I can confidently say that my graduate school experience has exceeded my expectations in every aspect. It wouldn't have been possible without the immense support and generosity of people and organizations that have played important roles in shaping my career from an early stage.

First and foremost, I would like to thank my Ph.D. advisor, Dr. Ishan Barman, for his constant support and encouragement throughout my doctoral studies. While Ishan's unique mentorship style allowed plenty of room for innovation and intellectual independence in selection of projects, his guidance and timely inputs have tremendously helped me build knowledge and be productive in new research areas. His optimism, diligence and intellect have been a constant source of inspiration during my tenure. In addition to directing the lab, Ishan is deeply invested in the professional development of his mentees. I hope to emulate his mentoring style in my future leadership roles.

As a doctoral student, I have also had the privilege of being mentored closely by Dr. Kristine Glunde and Dr. Narasimhan Rajaram. They have contributed significantly to my academic and professional growth during the course of my graduate studies. Kristine was instrumental in piquing my interest in biomedical applications of spectroscopy and helping me keep my studies rooted in clinical reality. Her patience and willingness to help me in obtaining necessary training, planning biological studies, and reshaping manuscripts and grants have tremendously helped in pushing our projects beyond the finish line. Narasimhan is an engineer with a strong training in biomedical applications of optical spectroscopy. By collaborating with him on several projects, I have learnt various aspects of interrogating biological systems using multimodal combination of optical methods. I would like to thank him for providing access to the cell lines and animal models produced in their lab to advance our collaborative projects.

I am grateful to Dr. Jeff Wang, Dr. Yun Chen and Dr. Kristine Glunde for serving as thesis readers and providing valuable suggestions for its improvement.

I will be indebted to my undergraduate research mentors at the Indian Institute of Technology Bombay – Dr. Amit Agrawal, Dr. Sudarshan Kumar and Dr. Jakub Kedzierski for sparking and sustaining my interest in academic research at an early age. They have trusted me with independent projects and necessary resources, an act that made me fearless in tackling new unexplored projects in areas that transcend traditional discipline boundaries. I would also like to thank my grandfather Mr. Govind Rajulu Paidi (late) and my many schoolteachers including Ms. K. Pushpavati, for generating an appreciation for mathematics.

The members of Barman Lab – both past and present – have played an important role in directing the course of my graduate studies. I have learned a great deal about spectroscopy and life in academia through our regular interactions and scientific discussions. The diversity of scientific backgrounds represented in the lab, ranging from physicists and chemists to engineers, has helped me gain a broader look and appreciation of the implications of our research beyond its immediate scope. I have worked closely with Soumik Siddhanta, Ming Li and Chi Zhang and pursued a variety of projects, many of which failed of course. But these explorations have definitely been enriching learning experiences. Outside lab, I

have spent amazing time exploring the United States with Soumik and Moumita on numerous expeditions during the last four years of my doctoral studies.

My academic success wouldn’t have been possible without the unwavering support of my friends – both at IIT Bombay and here at JHU. I would like to particularly thank Sayak Bhattacharya, Palash Agrawal, Parul Maheshwari, Hemanth Balaga, Deepak Venkat, Harsha Kolli, Amruta Bhavaraju, Vivek Nagal, Garima Shah, Dhananjay Sethi and Preeti Bhattacharjee. As a roommate of five years, Sayak has been an influential constant in my otherwise fast-paced and everchanging graduate school life. Palash, Hemanth, Deepak and Harsha continue to cheer me and help me navigate the ups and downs of my professional and personal life. I would also like to thank Parul for the much-needed distractions while writing this dissertation.

I would like to acknowledge the support of all the funding sources that have supported and nurtured my research over the years. Most notably, I would like to thank the Society of Laboratory Automation and Screening for providing me the SLAS Graduate Education Fellowship Grant and Tony B. Academic Travel Awards. I am also grateful to research funding from Johns Hopkins University, the National Institutes of Health, and Medimmune, LLC for supporting my research.

Finally, I would like to express my immense gratitude to the contributions made by my family. Since my childhood, my brother, Venkatesh Paidi, continues to challenge my aspirations and helps me keep my actions in sync with my goals. My parents, Simhachalam Paidi and Satyavati Paidi, have always encouraged us to dream big and fearlessly chase our ambitions. Their dedication to provide us

with the best environment for academic growth and the sacrifices they have made in the process cannot be put into words. By example, they have been successful in imparting in us, the value of hard work, integrity and compassion.

Dedication

To my teachers, family, and friends

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction to biomedical Raman spectroscopy

The diagnostic tools currently used in clinical practice largely rely on subjective evaluation of morphological manifestation of diseases. However, biochemical modifications often precede morphological changes in a variety of pathological transformations including cancer. A majority of laboratory techniques currently used to analyze the chemical make-up of biological samples are either too labor intensive or require operator expertise, which makes their clinical translation challenging. Powered by excellent chemical specificity and lack of extensive sample preparation requirements, Raman spectroscopy has emerged as an attractive solution for label-free characterization of biological systems both in vivo and ex-vivo/in vitro settings in a non-invasive or minimally invasive manner.

1.1 Raman spectroscopy

Raman spectroscopy, an inelastic optical scattering technique, was first discovered in 1928 by Sir C. V. Raman [1]. When monochromatic light is scattered by molecules, the elastically scattered photons are accompanied by lower or higher frequency photons (inelastically scattered light) due to changes in the vibrational states of the molecules under investigation. While a majority of photons undergo elastic scattering, a very small proportion lose energy (Stokes shift) and

---

This chapter is partially adapted from the published encyclopedia article: Paidi SK, Pandey R and Barman I. “Medical applications of Raman spectroscopy”, Encyclopedia of Analytical Chemistry, 1-21, 2020.

an even smaller proportion gain energy (anti-Stokes shift) at room temperature depending on the changes in vibrational states of the molecules (**Fig. 1.1**) [2, 3]. Most biological and biomedical investigation probe Stokes Raman scattered photons due to their relative abundance at room temperature. Consequently, biological Raman spectra in this region show strong fluorescence background due to excitation of endogenous chromophores in the samples. Since Raman spectroscopy doesn't require excitation of molecules to the stationary electronic states, a range of frequencies from deep UV to NIR region can be used for excitation depending on other considerations such as suppression of fluorescence, minimizing photodamage or maximizing penetration depth. This allows successful chemical structural fingerprinting of a variety of molecules independent of the excitation wavelength.

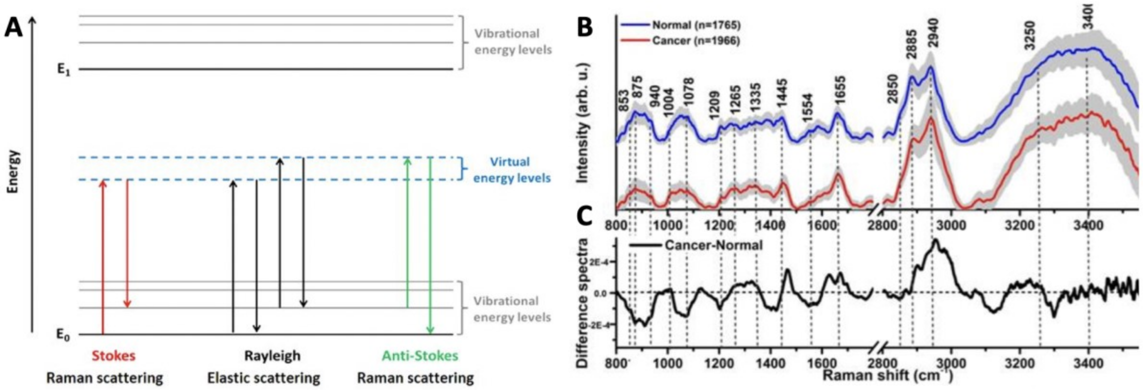


Figure 1.1: Biomedical Raman spectroscopy. (A) The energy level transitions for Stokes and anti-Stokes Raman spectroscopy are shown in comparison with Rayleigh scattering on a Jablonski diagram. Reproduced from ref. 2 (Licensed under CC BY SA 4.0). **(B)** Mean and standard deviation of Raman spectra obtained from normal and cancerous tissue sites in 95 patients during clinical endoscopy. Reproduced from ref. 3 (Licensed under CC BY NC 4.0). **(C)** The difference Raman spectrum (i.e., cancer - normal) shows disease-specific features. Reproduced from ref 3.

For several decades since its discovery, Raman spectroscopy was largely confined to the analysis of simple systems like pure liquids and gases. However, with the advent of powerful, stable and tunable lasers and sensitive multi-channel detectors in the past couple of decades, Raman spectroscopy has emerged as a powerful tool for analysis of complex biological systems like cells and tissues. Some of the early biochemical characterization of small and macro- molecules using Raman spectroscopy laid the foundations of future medical applications [4]. Important investigations of amino acids, proteins and enzymes in solid and aqueous states by Lord, Koenig, Yu and others revealed important insights about their secondary structures [5-7]. Similar studies conducted by Lord, Thomas, Hirano and others characterized the Raman bands of nucleotides and nucleic acids [6, 8-10]. Early Raman study of calcified tissues such as bone and collagen by Walton and others revealed important Raman bands characteristic of the constituents such as calcium phosphate, inorganic carbonate, proline and hydroxyproline [11] . Later studies of physiological and pathological calcification observed during bone remodeling and breast cancer have benefited from these Raman bands. Other studies around this time utilized the strong resonance enhancement offered by the substances such as hemoglobin, cytochrome c and β-carotene due to overlap of laser line with the electronic transitions [4]. Motivated by these observations, Raman studies of normal and diseased human tissues by several research groups soon followed in the 1990s. Several pioneering medical applications were focused on cancers of breast and brain, diabetic retinopathy, and mineralization, among others [12-17]. The ability to probe molecules in their native physiological states, negligible interference from water in biological samples, use of non-ionizing radiation and ease of integration with other techniques makes Raman spectroscopy particularly suitable for noninvasive characterization of biological samples *in vivo* [18, 19]. Medical applications ranging from disease diagnosis to metastasis assessment, chemotherapeutic monitoring to non-invasive glucose sensing, have been demonstrated using this technique [20-27].

Due to the use of visible-near infrared light, the clinical translation of Raman spectroscopy for medical applications is impeded by challenges such as poor sensitivity, broad fluorescence background, and poor penetration depth. To tackle these challenges, several variants such as surface enhanced Raman spectroscopy (SERS), spatially offset Raman spectroscopy (SORS), resonance Raman spectroscopy, coherent anti-Stokes Raman spectroscopy (CARS), and stimulated Raman Spectroscopy (SRS) have been developed and employed to enhance sensitivity and selectivity of Raman measurements [27, 28]. In addition, the developments in fiber-optics probes and advances in nanotechnology and computation have been vital for the recent surge in medical applications of Raman spectroscopy and its variants. In this chapter, we will discuss a selection of medical applications that have harnessed the advantages offered by Raman spectroscopy. In the first part, important aspects of instrumentation used in typical Raman spectroscopy applications will be briefly reviewed. This section will be followed by a brief summary of popular methods that are employed at each step in the data analysis workflow. These include background removal, dimensionality reduction and classification methods that are routinely used to create diagnostic frameworks based on Raman spectral data. Next, some variants of Raman spectroscopy that have been explored for medical applications are discussed to present the key advantages they provide over spontaneous Raman. In the final section, some representative studies demonstrating the feasibility of Raman spectroscopy in a wide range of pathologies and their key findings have been discussed.

1.2 Instrumentation

1.2.1 Illumination

The complexity of biological media requires monochromatic and narrow line-width light sources and sensitive detectors for obtaining interpretable Raman spectra and simultaneously minimizing laser-induced photodamage (**Fig. 1.2**) [29]. Therefore, the medical applications of Raman spectroscopy were made feasible by the developments and advancements in laser technology. Today lasers spanning from deep-UV to near-infrared (NIR) with small form factor are routinely used for Raman spectroscopy. The lasers in the UV region are particularly suited for resonance Raman spectroscopy applications due to the matching of their energy with the electronic transitions and the existence of an autofluorescence-free window in the deep UV region. However, they are not preferred for clinical Raman spectroscopy applications due to their propensity to cause photodamage. The lasers on the other end of the spectrum, NIR diode lasers (e.g. at 785 nm and 830 nm) offer significantly lower sensitivity due to the inverse dependence of scattering $(1/\lambda)^4$ on wavelength. Despite such disadvantage, existence of near absorption-free spectral window makes this region well-suited for tissue analysis by minimizing autofluorescence signal and maximizing the depth of penetration.

The lasers in the visible region offer higher Raman sensitivity than NIR sources while minimizing the laser-induced photodamage in comparison to UV sources for interrogation of systems such as single live cells, that are void of major biological chromophores. The compact assembly of modern semiconductor lasers allows for seamless integration of multiple lines spanning the spectrum from UV to NIR (and convenient switching between them) into laboratory Raman system. However, the portable clinical instruments built for endoscopy and surgical guidance often rely on a single NIR laser for tissue spectral acquisition.

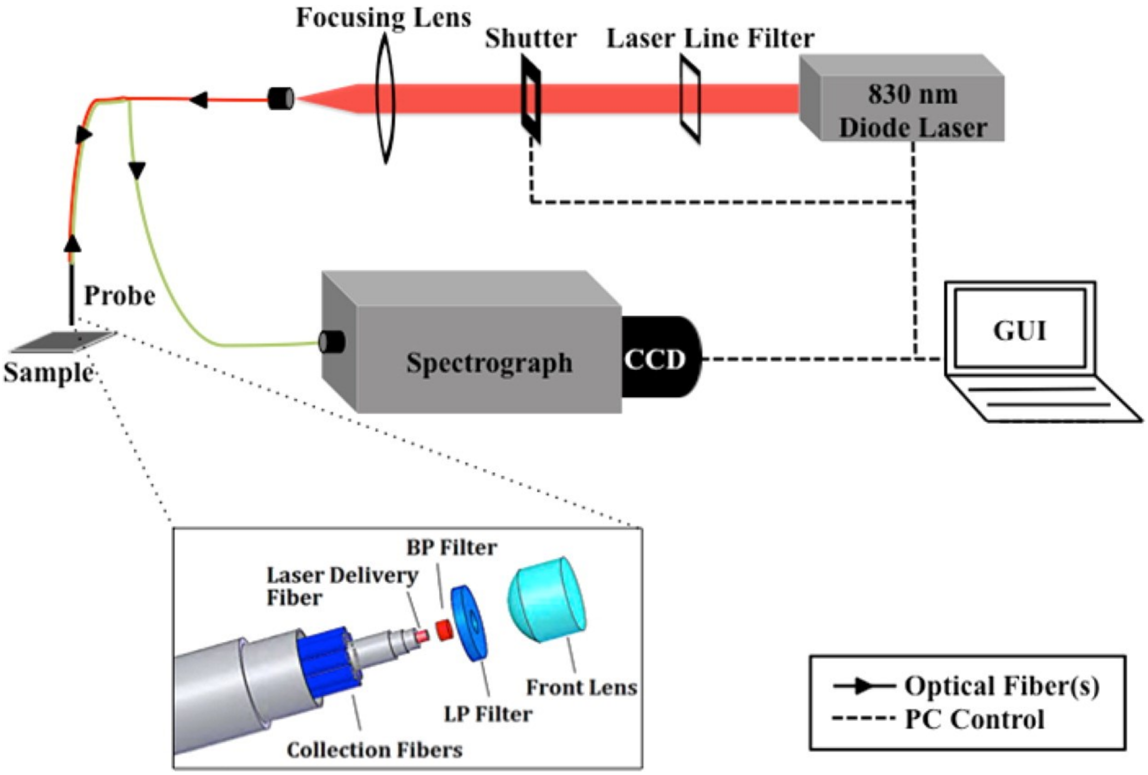


Figure 1.2: Schematic of a typical Raman spectroscopy system used for medical applications. Reprinted with permission from ref. 29. Copyright 2016 American Chemical Society.

1.2.2 Detection

The efficient collection of Raman scattered photons is very crucial for maximizing the throughput at minimal laser power and spectral acquisition time. The first critical step towards this is the rejection of elastically scattered light, which is typically accomplished by the use of edge or notch filters. Secondly, the decomposition of inelastically scattered light into constituent wavelengths (wavenumbers in Raman literature) is accomplished using diffraction gratings in most modern medical applications. The grating characteristics such as groove density (g/mm), focal length and groove profile play an important role in determining the spectral resolution of a Raman instrument. Typically, a single grating is sufficient to obtain spectra covering the fingerprint region at a satisfactory spectral resolution for clinical applications. Finally, the quantification of intensity at different wavenumbers is accomplished by focusing the dispersed radiation on a thermoelectrically cooled CCD. The CCDs are particularly suited for medical Raman spectroscopic applications in UV, visible and NIR regions due to their high quantum efficiency, low dark noise due to thermoelectrical cooling, and multichannel operation for the acquisition of complete spectrum in a single measurement. Furthermore, their compact architecture facilitates seamless integration of the resultant Raman spectroscopy systems with clinical workflow.

1.2.3 Raman probes and handheld devices

The effective delivery of laser and collection of backscattered Raman photons for disease diagnosis and surgical margin assessment has been made possible by developments in fiber optics technology [30]. Some of the key considerations in the design of Raman probes include maximizing collection efficiency of Raman signal and minimizing background signal stemming from fiber material itself. Most clinical applications employ optical fibers made up of fused silica glass doped with index-raising materials for increasing the refractive index of the core for obtaining total internal reflection. A significant proportion of probes also employ hollow core waveguides, which are hollow glass tubes with their interior coated with metals for reflection. While they eliminate the glass fiber background and resultant shot noise, they are associated with much higher losses and lesser flexibility. In the construction of Raman probes, multiple such fibers with designated delivery and collection roles are employed along with other optical elements such as filters and lenses. By selecting appropriate geometry, optical elements and materials, the probe characteristics such as throughput and depth selectivity can be modulated to suit specific applications (e.g. probing surface vs. subsurface lesions). The optical fiber-based Raman probes have been employed for various endoscopic diagnosis applications. The miniaturization of instrumentation is critical for further clinical acceptance of the technology and its translation from laboratory bench to bedside. While several handheld Raman detection systems have been designed and implemented for field applications in industrial settings, only few have been tested for clinical applications [31]. Improvements in the sensitivity of completely handheld Raman systems can be very beneficial for timely detection of disease in accessible organs such as skin, cervix and oral cavity, even in low resource settings. The handheld Raman detectors can be integrated with the fiber optic probes for delivery and collection of light from inaccessible regions such as breast, stomach and colon.

1.3 Multivariate data analysis

Raman spectra encode a wealth of molecular information stemming from the vibrational modes in a specimen. However, maximal utilization of the complex spectra obtained from biological samples that are composed of a large number of components requires systematic analysis of spectral datasets. Typically, Raman data analysis workflow includes steps such as preprocessing to remove spectral interferents, dimensionality reduction, and classification/regression. In this section, a selection of the most common data analysis techniques used in recent medical applications will be discussed.

1.3.1 Preprocessing

Background subtraction: Due to excitation of intrinsic fluorophores in the biological media [32, 33], the Raman spectra in the Stokes region are accompanied by a broad fluorescence background. Therefore, direct observation of weak Raman features amidst such overwhelming background is often not feasible. To tackle this, a variety of background removal methods have been reported [34]. The most common approach involves fitting a polynomial (e.g. fifth order) mimicking the broad background and subtracting it from the spectrum to uncover the weaker Raman features buried in it. Another method called asymmetric least squares that is used for background estimation and simultaneous smoothing relies on iterative minimization of the sum of the squared difference between the raw and estimated background signals using ordinary least squares with a bias towards positive residuals. In frequency domain methods such as those based on Fourier and wavelet transformations, the slowly-varying background can be distinguished from sharp Raman features and removed by appropriate frequency thresholding. These methods also allow smoothing by identifying and eliminating the contribution of high-frequency noise components. While these methods help uncover the latent Raman features, caution must be exercised to ensure that subtle features of diagnostic importance are not lost due to background subtraction. Studies have shown that in the presence of uniform background across the entire spectral dataset, the downstream classification algorithms yield very similar results for spectra with and without background removal [29, 35].

Smoothing: The weak Raman features stemming from biological specimen are often accompanied by noise and cosmic rays. Therefore, to better visualize Raman peaks and distinguish them from noise, different smoothing methods are routinely employed in the data analysis pipeline, often after background subtraction. The most commonly implemented denoising technique, Savitzky-Golay method involves locally fitting a polynomial of specified order within a moving window of fixed size. As a result, the technique is highly sensitive to the choice of polynomial order and window size because choosing a large window size can lead to over-smoothing and loss of important Raman features. Similarly, choosing a higher order polynomial in a small window can render the method ineffective due to overfitting and retention of noise. The other smoothing methods such as median filtering, wavelet transform, Gaussian filtering and methods based on rejection of high frequency components have been used for biomedical Raman data.

Normalization: Due to factors such as laser intensity fluctuations, differences in probe-sample distance and intensity fluctuations introduced by background subtraction, comparison of Raman spectra belonging to different samples acquired in non-uniform conditions can produce misleading observations. Therefore, there is a need to employ a suitable method for standardizing the spectra prior to analysis. This is often tackled by using intensity normalization using one of the many available techniques at one or more stages of Raman data analysis [34]. The most commonly employed normalization techniques for biomedical Raman data include peak normalization, vector normalization and min-max normalization. Traditionally, normalization of Raman spectra based on the intensity of a prominent peak stemming from an invariant component in the system-as an internal standard-was employed for comparison of the remaining peaks in a multi-component system. However, this method is not suitable for biological systems where identification of such an invariant component is difficult or where peak shifts are important indicators (e.g. evolution of protein structure). As a result, approaches such as min-max and vector normalization have been extensively used in the recent studies that set reasonable upper and lower bounds on the spectral datasets. In min-max normalization method, the maximum and minimum intensity levels of the spectra are made 0 and 1, respectively and the remaining intensities are scaled accordingly. However, in vector normalization, each spectrum is scaled to make its norm (typically Euclidean norm) unity.

1.3.2 Dimensionality reduction

Typically, Raman spectra used for medical applications, in the fingerprint or high wavenumber region obtained at an appropriate spectral resolution, have hundreds of dimensions (wavenumbers). This high-dimensionality, coupled with the availability of a limited number of spectra due to long acquisition times makes careful dimensionality reduction important for avoiding the ‘curse of dimensionality’ and improving the classification accuracy. Principal component analysis (PCA) is the most commonly employed technique to achieve dimensionality reduction by projecting the spectral dataset onto an orthogonal set of basis vectors in the order of the variance in the data they explain. Such condensation of important spectral information into fewer dimensions enables selective rejection of contributions from broad background and noise in the spectra as they can be easily identified from the obtained set of principal components. In addition to making data visualization simpler, the low-dimensional data thus obtained also enhances the robustness and computational simplicity of the downstream analysis such as classification and regression. The graphical representation of projection scores as 2D and 3D scatter plots or multidimensional radial visualization plots are often used for qualitative assessment of clustering behavior or other consistent patterns in the biological Raman datasets.

In addition to PCA, several other similar dimensionality reduction transformations such as independent component analysis (ICA), linear discriminant analysis (LDA) and partial least squares (PLS) have been used in the Raman literature to suit specific applications. Recently, multivariate curve resolution alternating least squares (MCR-ALS) has emerged as an important tool for identification of key spectral constituents that resemble dominant pure components in multicomponent mixtures [36]. Along with dimensionality reduction, the use of positivity constraints in the determination of MCR components allows for a direct comparison of the projection scores to gain important insights into the chemical abundance of the associated pure components across the samples. Therefore, the technique helps unbiased recovery of putative molecular markers from a biological specimen (as opposed to only spectral markers) that can serve as targets for traditional bioanalytical techniques.

1.3.3 Classification techniques

To fully exploit the chemical specificity of Raman spectroscopy and retrieve important information of diagnostic significance from subtle spectral differences in the biological samples, several supervised and unsupervised classification algorithms have been reported for successful diagnosis and monitoring of closely related pathological states. In this section, the merits and limitations of the most widely used techniques will be discussed.

Clustering methods: Clustering techniques provide an excellent opportunity for identification of intrinsic groupings and patterns in multivariate datasets, particularly in the absence of class labels. The two most popularly used methods in medical applications of Raman spectroscopy include k-means cluster analysis (KMCA) and hierarchical cluster analysis (HCA). In KMCA, based on the prescribed number of estimated classes and their centroids, the dataset is iteratively categorized into multiple clusters such that the distance between their centroids is maximized. While KMCA is simple and elegant for datasets with distinct classes, it is not very suitable if there exists a hierarchy within the classes. To tackle such datasets, HCA performs binary splits of the dataset progressively at each level via either a top-down or a bottom-up approach in order to create a dendrogram. In the top-down approach, the entire dataset is split into two clusters and the resultant clusters are then divided into two more clusters and so on until the clusters become too small or similar to perform further splits. Similarly, in the bottom-up approach, the individual instances are treated as distinct clusters and combined progressively based on their similarity to finally merge into one large cluster (i.e. the complete dataset). Due to the availability of the entire hierarchically organized dendrogram, HCA doesn’t require the knowledge of the number of classes in the dataset. The ability to use these clustering methods before or after dimensionality reduction can help identification of key mediators of class separation in spectral datasets.

Linear supervised classification: Several studies have successfully modeled Raman spectra of complex mixtures as a linear combination of the constituent spectra. Exploiting such linear relationships, several classification techniques that attempt to learn a linear plane separating the classes of interest have been adopted in biomedical applications of Raman spectroscopy. Linear discriminant analysis (LDA) and its variants are the most widely used methods for distinguishing closely related pathological states. The key assumptions about the data that allow the application of LDA include multivariate normality of each variable and equal covariance for all the classes. The lack of a sufficient number of instances compared to the dimensionality of spectra often encountered in Raman datasets makes direct application of LDA challenging due to the potential occurrence of non-full rank estimates of covariance matrices. This problem is often avoided by incorporating an appropriate dimensionality reduction method such as PCA to project the spectral dataset onto a low-dimensional space prior to applying LDA. Another widely used method in the biomedical Raman spectroscopy applications, partial least squares discriminant analysis (PLS-DA), attempts to find a discriminatory hyperplane by projecting both the spectra and their labels onto new spaces such that the covariance between the resultant matrices is maximal [37]. While the PLS loadings (unlike PCA loadings) are not easily interpretable, PLS-DA is particularly suited for spectral datasets with fewer instances than variables. However, under specific circumstances, the method has been shown to have similar performance as LDA and Euclidean distance to centroids, along with the associated disadvantages [37].

Non-linear supervised classification: While the linear classification methods are suitable for linearly separable datasets, a significant proportion of tasks encountered in the diagnosis of closely related and heterogenous pathological states using Raman spectroscopy demand robust modeling of non-linear class boundaries. To tackle such problems, support vector machines (SVM) have been frequently used in conjunction with appropriate kernel functions. SVM are used to identify a maximum margin class-separating hyperplane in the training dataset (**Fig. 1.3**). The use of appropriate kernel function allows building classification models for datasets that are not linearly separable by projecting them into high dimensional space and finding a suitable hyperplane in the transformed space. However, the obtained model, weights and parameters are not easily interpretable. Various implementations of SVM have been proposed in the literature along with a variety of optimization algorithms for classification in single, binary and multi-class settings. Appropriate selection of parameters for a given implementation is very critical to obtain high classification accuracy and avoid overfitting. LIBSVM, a very popular SVM library provides a powerful platform for implementing SVM and parameter selection in a variety of applications [38].

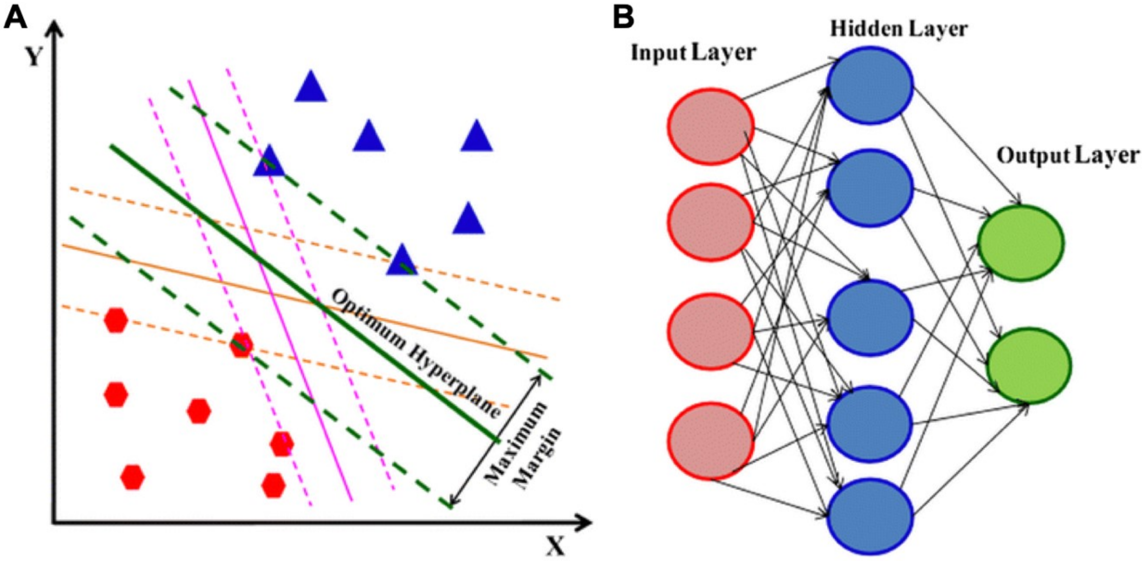


Figure 1.3: Multivariate data analysis. (**A**) Schematic representation of support vector machines showing identification of optical hyperplane for maximum margin between the classes. (**B**) Schematic representation of an artificial neural network with a single hidden layer showing the flow of data from input to output layer. Adapted from ref. 34 (Licensed under CC BY SA 4.0).

Artificial neural networks (ANN) is another powerful tool that has been widely used for modeling non-linear class boundaries in Raman spectroscopy datasets. The ANN classification, motivated by biological neural networks, involves modeling the relationship between training spectral dataset and class labels by learning the weights of inputs at each of the multiple transformations incorporated as a network of hidden layers between input and output layers (**Fig. 1.3**). Various methods such as backpropagation that iteratively learns and adjusts the weights of the network by minimizing the training error have made ANN fast and efficient. Due to the ability to model a wide range of complex functions, ANN with a large number of hidden layers can result in overfitting of the data for small datasets if caution is not exercised [39]. Additionally, the hidden layers serve as a black box and fail to identify the key contributors of the differences between classes of interest (e.g. benign vs. malignant cancers). However, the hardware developments enabling faster acquisition of large Raman datasets and emerging interest in deep learning are expected to make ANN and its variants more attractive.

Towards big data analysis: Traditionally, Raman spectroscopic datasets have been relatively small due to the low throughput and lack of automation in spectral data collection. However, recent developments in automation and integration of Raman spectroscopy and its faster variants that utilize surface enhancement (SERS) and non-linear processes (CARS and SRS) with advanced microscopes have enabled collection of large Raman datasets from a variety of biological media such as cells, tissues, and body fluids. The successful acquisition of large spectral datasets using these technologies, especially 2D and 3D Raman maps will enable researchers to translate the recent success in image classification using deep learning frameworks. The standardization of data acquisition protocols for spectral acquisition of cells and tissues for each class of diseases will allow us to build, share and continuously expand the Raman datasets to capture inter-patient heterogeneity, train robust deep learning models and translate other emerging big data tools.

1.4 Variants of Raman spectroscopy

In addition to spontaneous Raman spectroscopy, several variants have been demonstrated for medical applications where attributes such as higher sensitivity, depth selectivity, specific targeting and resonance enhancement are desirable. In this section, some prominent variants of Raman spectroscopy with potential for clinical translation will be briefly discussed.

1.4.1 Surface enhanced Raman spectroscopy (SERS)

A significant enhancement of electromagnetic field is achieved close to the metal surfaces due to excitation of localized surface plasmons. Surface enhanced Raman spectroscopy exploits this phenomenon for selective enhancement of Raman signal from molecules adsorbed on the metal surfaces by up to 11 orders of magnitude [40]. Metallic nanostructures made up of gold and silver have been used for SERS applications owing to their strong surface plasmon resonance in visible and NIR regions, where most Raman laser lines lie. Various nanostructures such as spheres, cubes, stars, rods and wires have been extensively studied for obtaining maximal SERS enhancement for a variety of applications in biology and medicine. Most studies utilize SERS enhancement to boost the detection sensitivity of low concentration analytes in one of the following two configurations. The label-free SERS or direct detection involves bringing the analyte close to the nanostructures for maximal enhancement of their intrinsic Raman signatures. This configuration has been used for studying biofluids, tissues and cells for identification of disease biomarkers. While untargeted enhancement of all the constituents of biological media improves the sensitivity of detection (and reduction in laser exposure), it fails to exploit known markers for increased specificity via targeting. Therefore, targeted applications, also known as indirect detection, employ nanoprobes composed of Raman active reporter molecules tightly packed with the metal nanostructures. The surface of such nanoprobes can be modified to selectively bind with the analyte of interest for its indirect quantification by measuring the signal from the Raman reporter molecules. However, the widespread use of SERS tags for *in vivo* applications is impeded by concerns about toxicity of the used nanoparticles. Several constructs involving the use of a thin protective shell gold, silica or polymers (e.g. polyethylene glycol) have been proposed to improve the biocompatibility of SERS nanoparticles.

1.4.2 Resonance Raman spectroscopy

The weak Raman signals can be significantly enhanced when the excitation frequency is resonant with one of the electronic absorption bands [41]. Such enhancement enables preferential enhancement of vibrational modes associated with chromophores without interference from other components of complex mixtures in which they are present. For example, such resonant enhancement allows the study of specific parts of large protein molecules, e.g. the heme region in myoglobin, without affecting the signal from other regions. However, since most lasers currently used in Raman spectroscopy are not easily tunable, only a few modes whose absorption bands match the wavelengths of the available laser lines can be conveniently probed. Additionally, due to the use of lasers in the UV and visible regions and absorption by the chromophores of interest, the resonance Raman spectra are almost always associated with a strong fluorescence background. Fortunately, strategies such as using shorter wavelengths into the deep-UV region and time-gating for selectively rejecting the fluorescent scattered photons have shown promise for alleviating the strong fluorescence background and associated shot noise in resonance Raman measurements. A range of applications focused on the spectroscopic measurement study of molecules such as nucleic acids, proteins and metalloproteins in their native states as well as in complex biological matrices, have utilized resonance Raman spectroscopy.

1.4.3 Spatially offset Raman spectroscopy (SORS)

The poor penetration depth provided by visible and NIR lasers used for Raman spectroscopy prevents interrogation of deeper layers of tissue. To obtain Raman signal from the deeper regions, spatially offset Raman spectroscopy employing a configuration that involves spatial separation of the illumination source from the detection zone has been used (**Fig. 1.4**) [28]. Such illumination-collection separation provides a route for preferential retrieval of spectra from deeper layers of diffusely scattering media by combining measurements at multiple offsets and multivariate data analysis methods. Various source-detector separation configurations involving single or multiple zones of illumination as well as detection using fiber optics have been demonstrated for biomedical and pharmaceutical applications. One such configuration featuring central collection zone and a spatially offset illumination ring (using axicons), called inverse SORS is particularly suited for clinical applications due to the lower laser power density and multiple spatial offsets achieved by the use of ring illumination. While SORS has shown promise for applications such as breast cancer detection, tomographic imaging and blood glucose detection, its implementation is particularly challenging for highly absorbent and fluorescent samples.

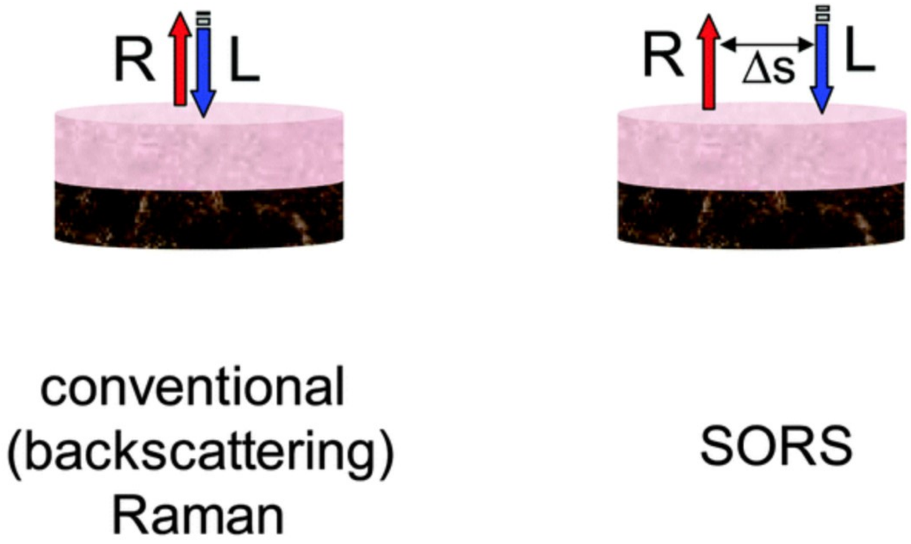


Figure 1.4: Schematic representation of the spatially offset Raman spectroscopy (SORS) principle in comparison to spontaneous Raman spectroscopy. R, L and Δs denote backscattered, laser beam and spatial offset, respectively. Adapted from ref. 28 (Licensed under CC BY SA 3.0).

1.4.4 Non-linear Raman spectroscopy

The non-linear techniques, particularly coherent anti-Stokes Raman spectroscopy (CARS) and stimulated Raman spectroscopy (SRS), have emerged as attractive variants to address low yield and strong fluorescence background encountered in biological species. Both the methods have been successfully integrated in microscopes for investigation of cells and tissues. In CARS, a coherent anti-Stokes signal is generated by the interaction of pump, Stokes and probe frequencies in a four-wave mixing process as shown in **Fig. 1.5** [42]**.** The key advantages provided by CARS over spontaneous Raman include the suppression of fluorescence, higher yield due to the coherent laser-like output, and the dependence of spectral resolution only on the excitation laser width. However, non-resonant background signal from the matrix can limit the specificity of the technique by obscuring the CARS signal at low concentrations. CARS microscopy has been utilized for visualization of lipids and other biomolecules in live cells by appropriately tuning the difference frequency between pump and stokes beams to match specific vibrational bands (e.g. C-H band in lipids) with high 3D spatial resolution. On the other hand, SRS microscopy uses pump and Stokes laser beams such that their frequency difference matches specific molecular vibrational frequency of interest to achieve significant enhancement of the associated Raman signal due to stimulated excitation without non-resonant background (**Fig. 1.5**) [42-45]. The linear dependence of SRS signal with the concentration of analyte makes it suitable for direct quantification of individual components in multicomponent mixtures. Emerging medical applications of SRS microscopy have used the higher sensitivity and specificity to achieve fast mapping of cells and tissues (**Fig. 1.6**) [43]. While, the inability to probe multiple Raman bands simultaneously used to be a major drawback of SRS microscopy, recent studies have shown multiplexing using a variety of methods [44, 45].

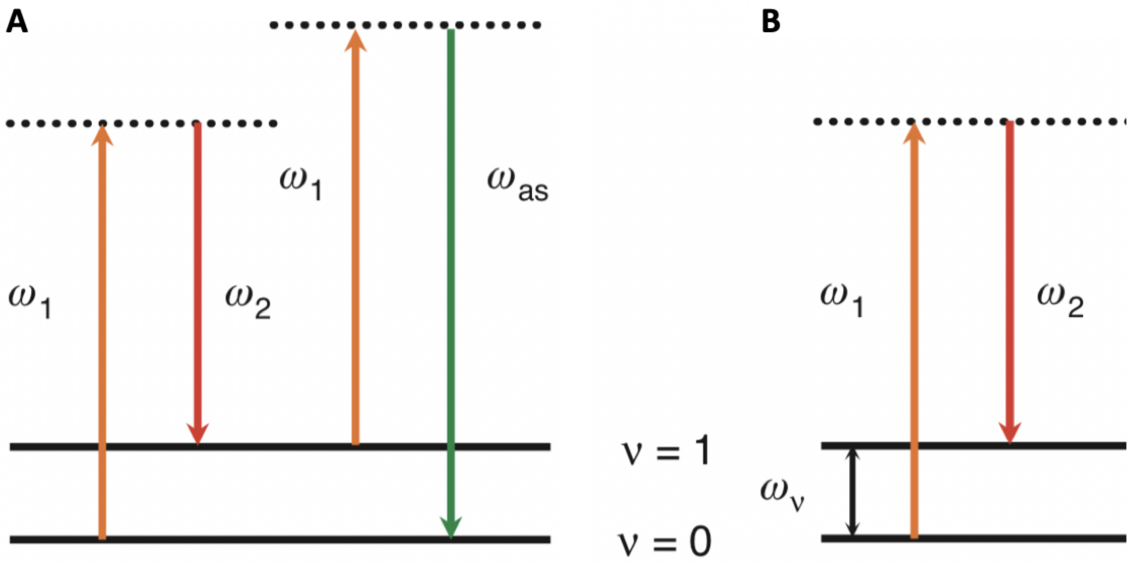


Figure 1.5: Jablonski diagrams. Jablonski diagrams for **(A)** CARS and (**B**) SRS processes are shown. Reproduced from ref. 42 (Licensed under CC BY SA 3.0).

1.5 Medical applications of Raman spectroscopy

In recent years, the excellent sensitivity and chemical specificity offered by Raman spectroscopy and its variants have been exploited for a variety of medical applications through clinical and pre-clinical studies. In this section, we will discuss some important applications in areas such as cancer diagnosis and chronic disease management.

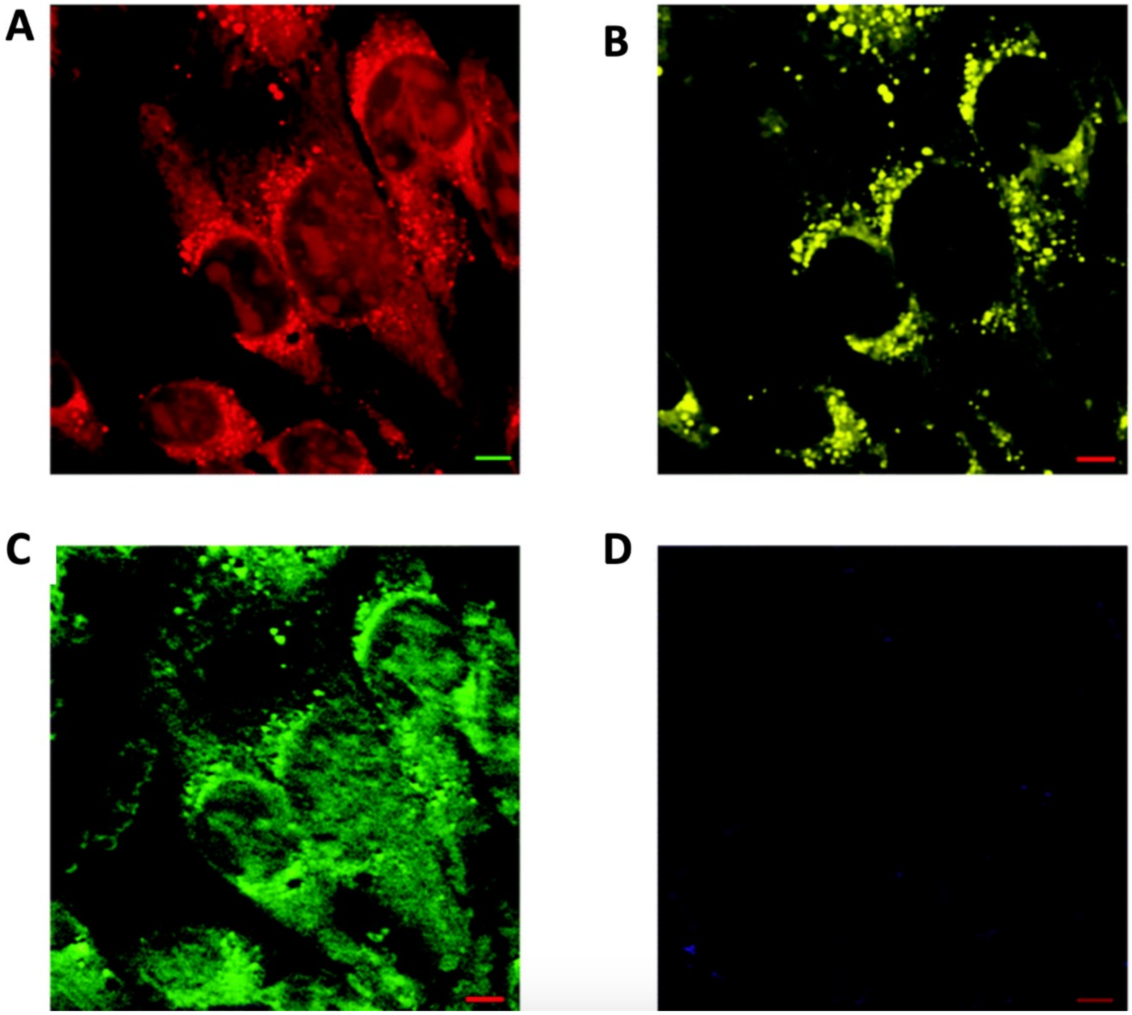


Figure 1.6: SRS microscopy of cells. Label free stimulated Raman scattering microscopy of fixed squamous cell carcinoma cells acquired at (**A**) 2953 cm^{-1} (CH_3), (**B**) 2850 cm^{-1} (CH_2, lipid), (**C**) 1655 cm^{-1} (C═O, amide-I), and (**D**) 1700 cm^{-1} (cellular silent region). Scale bars are 5 μm. Reproduced from ref. 43 (Licensed under CC BY SA 3.0).

1.5.1 Applications in cancer

Clinical Diagnostics: Breast cancer, one of the most common forms of cancer in women, has been extensively studied using Raman spectroscopy and its variants. Early investigations of breast cancer using Raman spectroscopy exploited high Raman scattering cross-section exhibited by mammary microcalcifications, which are frequently used to assess breast disease. The benign and malignant breast cancers have been accurately identified based on differences in Raman signatures of type I (composed of calcium hydroxyapatite) and type II (composed of calcium hydroxyapatite with carbonate substitution) mammary microcalcifications, respectively [46]. Raman spectroscopy has also been employed to discern a variety of breast lesions ranging from benign fibroadenoma to invasive ductal hyperplasia in samples that do not present microcalcifications, with accuracies upwards of 85% for the diseased classes [47]. The enhancement of surface Raman signal using silica-shell isolated nanoparticles for the same samples provided further improvement in accuracy (> 90%) for most classes. The use of multivariate data analysis tools like PCA provides important insights into the disease-specific compositional changes in the tumor microenvironment. Using controlled tissue constructs composed of calcium salts (hydroxyapatite, oxalate and carbonate-substituted hydroxyapatite) embedded in chicken tissue at varying depths, the efficacy of detecting microcalcifications from deeper subsurface regions using SORS was demonstrated for depths ranging from 2 mm to 10 mm [48]. SERS has been explored extensively for studying breast cancer cells and biofluids from patients by incorporating a variety of nanoparticles. A majority of

SERS studies for breast cancer cells have been conducted for targeted quantification of cell surface markers such as HER2 and EGFR, that are differentially expressed in cancer cells, via nanoprobes modified with relevant antibodies for preferential binding to the antigens expressed on their membranes. The SERS nanoprobes used in these studies can be readily adapted for targeted identification of cancer cells from other organs if appropriate biomarkers are known. *In vivo* targeting of cancer cells in tumor xenografts pegylated gold nanoparticles containing organic Raman reporter molecules has been demonstrated by targeting EGFR using single-chain variable fragment (ScFv)-antibody conjugation [49].

Among other accessible cancers, the diagnosis of skin cancer has also been an important application of Raman spectroscopy due to high incidence rates and sampling errors encountered during biopsy for diagnosis. Several groups have studied benign and malignant skin cancers both *in vitro* and *in vivo* for developing diagnostic algorithms for real-time noninvasive detection. In a comprehensive Raman spectroscopy study incorporating 453 patients presenting various types of melanomas, carcinomas, keratoses and nevi, principal component general discriminant analysis and partial least squares have been used for three different diagnostic classification tasks – skin cancer and precancers vs. benign lesions, melanomas from nonmelanoma lesions, and melanomas from seborrheic keratoses with area under the curve (AUC) values of over 0.82 for each [50]. Several applications in cervical cancer screening have been demonstrated based on Raman spectroscopy to detect cytopathological changes, HPV infection status (a key risk factor) as well as differences in protein, lipids and nucleic acid content between benign and malignant lesions [51]. Early *in vivo* study using fiber optic probes revealed an association between an increase in collagen, phospholipid and DNA signatures and progression of tissue from normal to high-grade squamous dysplasia [52].

The *in vivo* application of Raman spectroscopy for diagnosis of inaccessible cancers requires seamless integration of the fiber optic probes with endoscopy workflow. Motivated by *ex vivo* studies that demonstrated the potential for use of fiber-optic Raman probes for *in vivo* diagnosis, studies have focused on endoscopic detection of cancers and other abnormalities in esophagus, stomach and colon. The LDA classification model developed using *in vivo* esophageal Raman measurements obtained from 27 patients during clinical endoscopic examination of esophagus and biomolecular modeling yielded sensitivity and specificity of 97% and 95.2%, respectively for the diagnosis of esophageal cancer [53]. The biomolecular modeling revealed that neoplasia was associated with a decrease in actin, collagen, lipids and glycogen as well as an increase in DNA and histones concentration. *In vivo* studies in rats have showed the feasibility of using topically applied antibody-conjugated SERS nanoparticles for sensitive detection and quantification of cell surface markers of esophageal cancer [54]. Translation of such protocols to human patients are expected to enable identification of hormone receptor status along with disease diagnosis. *In vivo* Raman endoscopic examination of gastric tissue in 305 patients enabled the construction of a spectral database for developing a diagnostic algorithm based on PLS-DA for in-line detection of gastric cancer in independent patient samples with a predictive accuracy of *ca.* 80% [55]. In the colon, various polyps and adenocarcinomas have been studied both *in vivo* and *ex vivo* using fiber-optic Raman probes that are compatible with conventional colonoscopes [56]. Additional advancements such as utilizing high wavenumber spectral regions for diagnosis have hinted at improving the capabilities of Raman spectroscopy in endoscopic applications.

Intra-operative surgical guidance: The lack of complex sample preparation steps and minimal perturbation to the tissue due to NIR lasers used in Raman spectroscopy makes it an attractive tool for real-time assessment of margins during tissue conserving surgeries. One of the first studies in this direction was conducted *in vivo* in 9 patients undergoing partial mastectomy breast surgery, where previously developed classification algorithms have yielded accurate classification of malignant among normal and benign tissues [57]. However, the residual subsurface tumors under surgical margins can result in recurrence of cancer. To circumvent this challenge, spatially offset Raman probes have been developed and validated for successful classification of positive and negative margins in frozen-thawed breast surgical samples *ex vivo* [58]. Another *ex vivo* study using breast surgical samples showed the potential for accelerating acquisition of Raman micro-spectroscopy data by employing autofluorescence imaging for selective sampling of spatial locations [59, 60]. Such multimodal approaches can significantly reduce the Raman data acquisition time and provide higher sensitivity for real-time detection of tumor margins during surgery. *In vivo* application of Raman spectroscopy in 38 bladder cancer patients during transurethral resection, sensitivity and specificity over 75% have been achieved using PCA/LDA classification analysis [61]. In brain cancer, another organ where conserving tissue during surgery is critical for retaining key functions and memory, a recent study employed 17 patients undergoing resection of grade 2 to 4 gliomas for *in vivo* identification of cancer margins composed of high and low density tumor cells using a fiber-optic probe with sensitivity and specificity higher than 90% [62].

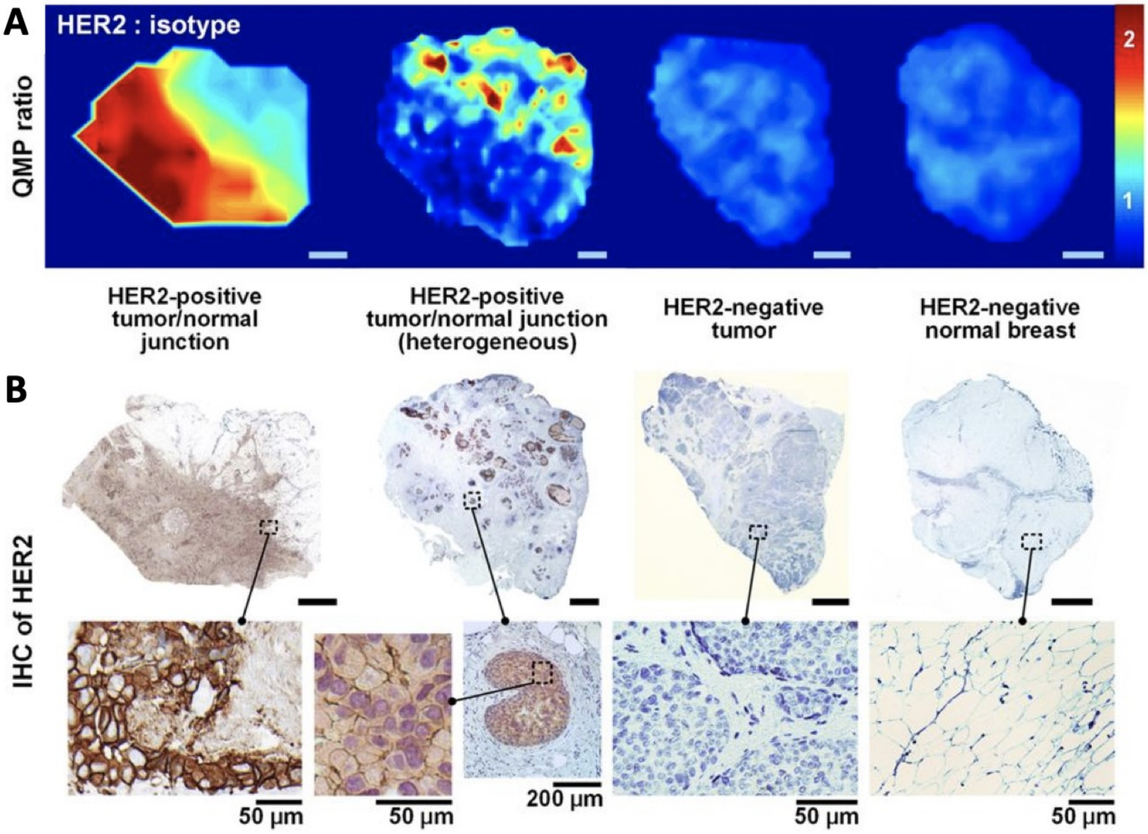


Figure 1.7: SERS imaging of tissues markers. (A) SERS imaging of two HER2-positive specimens containing both tumor and normal tissue regions and two HER2-negative specimens (one tumor and one normal tissue) by topical administration of a nanoparticle (NP) mixture of HER2-NPs and isotype-NPs for targeted identification of regions overexpressing HER2 receptor based on the ratio of bound NPs. **(B)** Immunohistochemical staining of the same samples with an anti-HER2 antibody. Unlabeled scale bars represent 2 mm. Adapted from ref. 63 (Licensed under CC BY 4.0).

Some early studies have also shown the feasibility of using SERS imaging of resected samples for intra-operative margin assessment and surgical guidance in clinic. The *in vitro* success in quantifying cell surface receptors that are overexpressed in cancer using targeted SERS nanoparticles has motivated similar studies for identification of margins by topical application of nanoparticles on surface of freshly excised tissue specimen. In a recent study, topical application of a multiplexed cocktail of SERS nanoparticles has resulted in quantification of molecular phenotype in terms of receptor status. Using a mixture of SERS nanoparticles functionalized with HER2 and isotype antibodies, tumor margins have been determined based on HER2 status in freshly excised tissue samples obtained at breast cancer lumpectomy (**Fig. 1.7**) [63]. Another similar study employed a similar approach for simultaneous quantification of HER2, ER, EGFR and CD44 in 57 clinical sample for detection of breast carcinoma with sensitivity and specificity of 89.3% and 92.1%, respectively [64]. A study targeted at understanding the biodistribution of ingested SERS nanoparticles in the body has revealed their potential to be used *in vivo* for topical application in oral cavity and GI tract without worrying about permeation into the blood stream or accumulation in vital organs [65].

Evaluation of metastatic disease: While Raman spectroscopy has been extensively used for establishing feasibility of clinical translation in cancer diagnostics, emerging studies have also focused on understanding the molecular mechanisms driving key characteristics of cancer and its metastatic progression. In mice harboring metastatic brain tumors induced by murine melanoma cells, Raman spectroscopy using a fiber-optic probe enabled the detection of cortical and subcortical tumor cell aggregates in the brain based on spectral features stemming from proteins, lipids, blood, water, bone, and melanin [66]. Similar mice studies of prostate cancer metastasis to bone using Raman spectroscopy revealed compositional changes in the bone such as an increase in carbonate substitution in hydroxyapatite and a decrease in collagen mineralization and mineral crystallinity (often correlated with physical properties of bone), that were independent of architectural alterations determined using microCT [67]. A recent study by our group showed that Raman spectroscopy was able to identify early biochemical changes in the composition of bones of mice bearing breast tumor xenografts, before the emergence of morphological indicators discernible by radiographic imaging [68]. The metastasis of breast cancer cells to lymphatic system in patients is routinely determined by sampling axillary lymph nodes. The Raman spectroscopic evaluation of 38 axillary lymph nodes (25 negative and 13 positive for metastatic spread) obtained from 20 patients undergoing breast cancer surgery were used to create a leave one node out cross validation routine based on PC-LDA that yielded sensitivity and specificity over 90% for prediction of normal and metastatic lymph nodes [69].

A large number of studies have also focused on employing Raman spectroscopy to study metastasis at cellular level by employing spontaneous Raman as well as labeled and label-free SERS using a variety of nanoparticles. A recent study established the feasibility of employing label-free Raman spectroscopy for classification of isogenic cell lines that show organotropism to brain, liver, lung and spine [70]. The findings were supported by metabolomic analyses that showed distinctions in the cell lines based on organotropism. However, identification of rare cells in complex matrices often requires the use of labeled SERS for obtaining the required sensitivity. A study aimed at detection of circulating tumor cells (CTCs) in human peripheral blood derived from patients with different stages of head and neck squamous cell carcinoma accomplished it by using EGF-conjugated SERS nanoparticles for targeting epidermal growth factor expressed by the CTCs [71]. They showed that the SERS intensity of the samples correlated with the number of CTCs present per mL of the sample. These studies show that Raman spectroscopy and its variants are suitably poised to tackling advanced challenges in cancer such as determination of metastatic risk and organotropism.

1.5.2 Applications beyond cancer

The medical applications of Raman spectroscopy are not limited to oncology. A significant amount of work has also focused on key areas of human health such as management of chronic diseases, diagnosis of infectious diseases, and prediction of risk of acute illness. Non-invasive glucose detection for diabetes monitoring has been one of the long-standing goals in Raman spectroscopy. Several studies have made progress towards accomplishing this goal by addressing several challenges such as sample-to-sample variability, autofluorescence due to endogenous chromophores, tissue turbidity, physiological lag between blood and interstitial fluid glucose levels, and non-linear spectra-concentration relationship (**Fig. 1.8**) [72]. For instance, turbidity correction by combining diffuse reflectance and Raman spectroscopy measurements under the assumption that the effects of turbidity manifest similarly in both the modalities yielded significant improvement in glucose concentration prediction. Similarly, the inaccuracies induced in the spectra-concentration calibration models due to the difference in glucose concentration between blood and interstitial fluid compartments have been mitigated by modeling the physiological glucose dynamics using mass transfer principles.

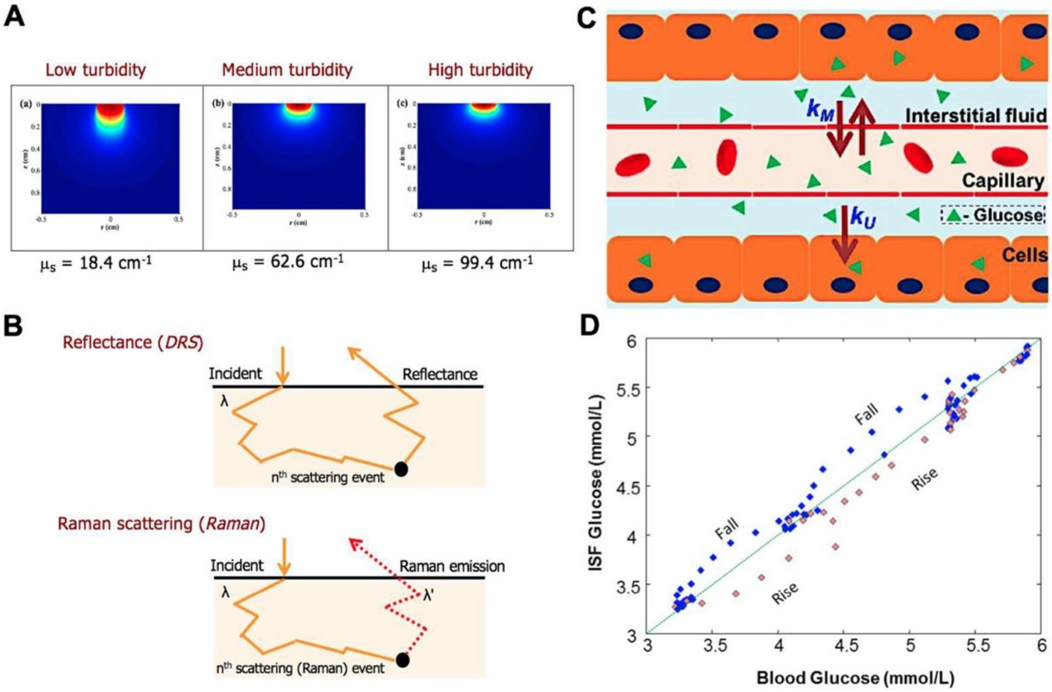


Figure 1.8: Non-invasive glucose monitoring using Raman spectroscopy. (A) Simulations of effects of tissue turbidity on sampling volume is shown for different scattering coefficients. **(B)** Schematic showing the similarity of photon-tissue interactions for diffusely reflected and Raman scattered light of same frequency. **(C)** Illustration of glucose diffusion between the blood and interstitial fluid (ISF) as well as cellular uptake that cause inaccuracies in calibration models built to predict blood glucose levels based on Raman measurements of ISF glucose. **(D)** The resultant lag between the glucose in blood and ISF compartments leads to differences in measurements during rapid changes (rise/fall) in glucose concentrations. Reprinted with permission from ref. 72. Copyright 2017 American Chemical Society.

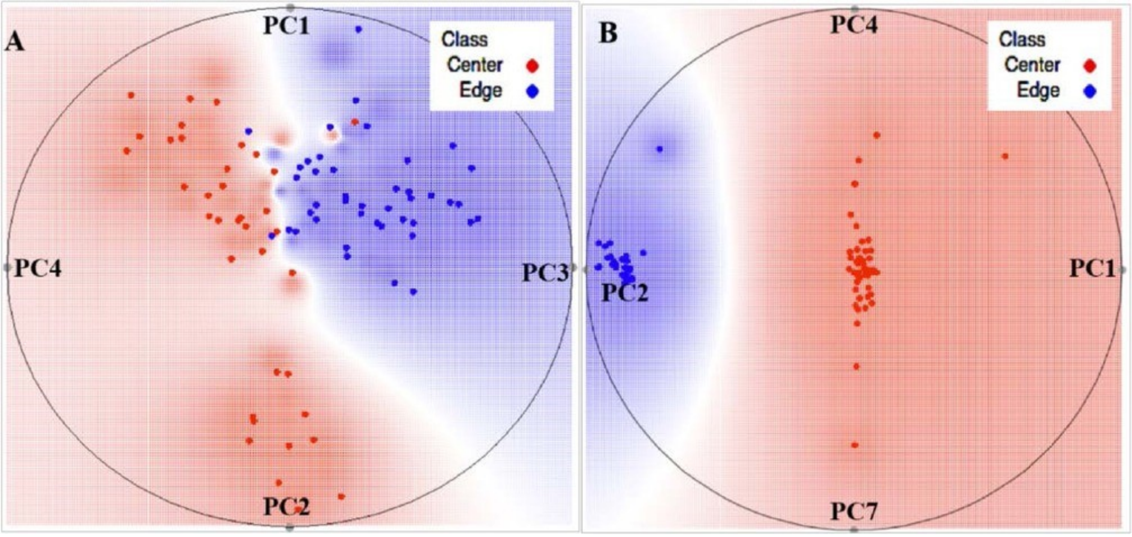


Figure 1.9: Identification of fungal growth zones using Raman spectroscopy. Radial visualization PC score plots derived from **(A)** spontaneous Raman and **(B)** label-free SERS spectra acquired from center and edge of the *P. indica* fungal cultures are shown. Reproduced with permission from ref. 74. Copyright 2017 Wiley-VCH Verlag GmbH & Co. KGaA.

The diagnosis of infectious disease agents is another important area where the ability of Raman spectroscopy to provide rapid diagnosis is immensely valuable. One of the early studies utilizing NIR confocal Raman micro-spectroscopy of common microorganisms on solid culture medium showed test dataset classification accuracy upwards of 80% for the combined spectral dataset composed of multi-day measurements from five types of bacterial microcolonies using a combination of PCA and LDA [73]. More recently, the study of different growth zones of fungal cultures by us using a combination of light sheet microscopy and Raman spectroscopy revealed that the biochemical basis of morphological differences across the growth zones were sufficient for accurate classification of the growth zones using PLS-DA [74]. By dispersing SERS nanoparticles on the surface of fungal culture further improved the classification accuracy without specific targeting (**Fig. 1.9**). However, the ability to identify microorganisms in complex matrices such as blood and other biofluids is critical for field translation. Direct identification of predominant bacterial species causing urinary tract infections (UTI) in urine samples collected from 10 patients was demonstrated by combining Raman micro-spectroscopy and SVM model trained on a spectral database built using 11 important UTI bacterial species, without the need of a cell culture step before acquisition of Raman spectra [75]. Similarly, Raman spectroscopy and PCA has been used to monitor changes in erythrocytes and plasma associated with progression of malaria by exploiting the resonant enhancement of hemoglobin and hemozoin [76]. Several other pathologies such as coronary atherosclerosis, middle ear diseases, osteoarthritis and dental abnormalities have been explored using Raman spectroscopy and its variants in conjunction with multivariate data analysis techniques [77-81]. For instance, successful classification of infracted and non-infracted tissue regions in myocardial in surgically excised ventricular myocardium has been recently shown for label-free evaluation of myocardial infraction (**Fig. 1.10**) [82].

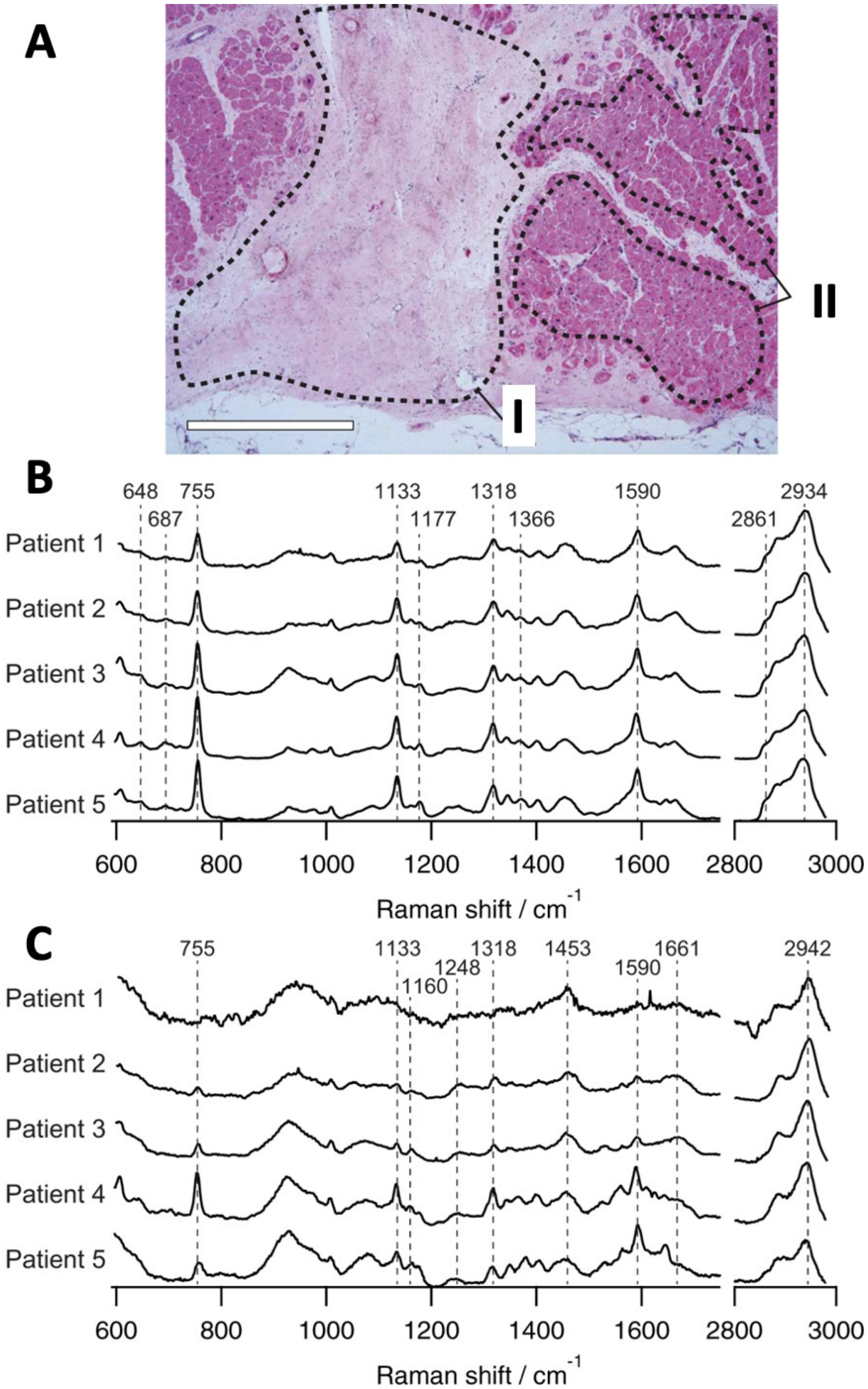


Figure 1.10: Detection of myocardial infarction using Raman spectroscopy. (A) A representative H&E-stained image showing the boundary between the infarcted (labeled I) and non-infracted (labeled II) regions of the heart tissue excised from human patients is shown. Representative mean Raman spectra belonging to **(B)** non-infarcted myocardium and **(C)** infarcted myocardium regions in the samples from each patient is shown. Scale bar represents 1 mm. Reproduced from ref. 82 (Licensed under CC BY SA 4.0).

1.6 Thesis outline

This dissertation aims to highlight our recent work that leverages the exquisite molecular specificity of Raman spectroscopy to study cancer progression and its response to therapy. Following the brief introduction to medical applications of Raman spectroscopy that is laid out in the present chapter, I first focus on our investigations into the utility and genetic basis of mammary microcalcifications as markers of primary breast cancer diagnosis. In the later chapters, I discuss our studies focused on the spectral characterization of pre-metastatic niches and response to radiation therapy, applications that go beyond primary disease diagnostics. The thesis is organized as follows:

In Chapter 2, we demonstrate the use of Au@SiO_2 shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS) for assessment of microcalcifications and distinguishing between normal breast tissues, fibroadenoma, atypical ductal hyperplasia, ductal carcinoma in situ (DCIS), and invasive ductal carcinoma (IDC) in human patient samples. We assess the effect of SHINERS signal enhancement on the accuracy of classification routine used for the identification of breast lesions with and without microcalcifications. By comparing the classification accuracies obtained in lesions with and without microcalcification using both SHINERS and traditional spontaneous Raman spectroscopy, we also examine the magnitude of relative improvement provided by the pursuit of mammary microcalcifications for diagnosis of breast cancer in current clinical practice.

In Chapter 3, we investigate the genetic basis of breast cancer microcalcifications and elucidate their role in metastasis. Guided by the gene expression analysis of metastatic and non-metastatic breast cancer cell lines, we show that among the genes previously implicated in the formation of microcalcifications, Osteopontin (OPN) varies the most between the two groups. We study the direct effect of OPN on the osteogenic cocktail-induced microcalcification formation and cell migration via stable shRNA silencing of OPN in MDA-MB-231 cells. Furthermore, we also compare the relative levels of OPN mRNA in parental MDA-MB-231 cells with those derived from circulation (CTC) and lungs (LM) of mice implanted with the parental cells, to verify correlation of OPN with cell migration.

In Chapter 4, we turn our attention to the metastatic disease. Here, we test the ability of label-free Raman spectroscopy to detect early changes in the future secondary sites of metastatic disease in advance of tumor cell arrival from the growing primary tumor. We use high metastatic MDA-MB-231 and low metastatic MCF-7 cells to grow orthotopic breast cancer xenograft for spontaneous dissemination from the mammary fat pad and potentially metastasize to the lungs. We map the lungs of the mice excised during the pre-metastatic phase and acquire spectral dataset for multivariate data analysis and identification of markers indicative of subtle biomolecular changes in the lung microenvironment as a result of the formation of pre-metastatic niches. We use histopathology and gene expression analysis to support the Raman features identified in the study.

In Chapter 5, we explore the possibility of monitoring response to radiation therapy at clinically relevant dosages and prospective prediction of responders using Raman spectroscopy. Here, we employ an isogenic matched model of radiation resistance consisting of A549 lung cancer cells and their resistant variant rA549 cells to grow tumors in mice to reliably study the differences due to sensitivity to radiation therapy. We also use tumors derived from sensitive and resistant head and neck cancer cells to delineate model-specific markers of response and sensitivity. Using multivariate curve resolution alternating least squares (MCR-ALS) and support vector machine (SVM) classification, we determine the putative markers of radiation response and sensitivity and build models for prospective prediction.

Finally, in Chapter 6, I summarize the key findings of the abovementioned studies and discuss some future research directions that can build on the recent developments in the area to facilitate the rapid translation of this promising technology to the clinic for disease diagnostics and beyond.

Bibliography

[1] C. V. Raman and K. S. Krishnan, A new type of secondary radiation, *Nature* **121**(3048), 1928, 501-502.

[2] K. J. Ember, M. A. Hoeve, S. L. McAughtrie, M. S. Bergholt, B. J. Dwyer, M. M. Stevens, et al., Raman spectroscopy and regenerative medicine: A review, *NPJ Regenerative Medicine* **2**(1), 2017, 12.

[3] K. Lin, W. Zheng, C. M. Lim and Z. Huang, Real-time in vivo diagnosis of nasopharyngeal carcinoma using rapid fiber-optic Raman spectroscopy, *Theranostics* **7**(14), 2017, 3517.

[4] F. S. Parker, Biochemical applications of infrared and Raman spectroscopy, *Appl Spectrosc* **29**(2), 1975, 129-147.

[5] R. C. Lord, Laser-Raman spectroscopy of biological macromolecules, *XXIIIrd International Congress of Pure and Applied Chemistry* 1971, 179-191.

[6] J. L. Koenig, Raman spectroscopy of biological molecules: A review, *Journal of Polymer Science: Macromolecular Reviews* **6**(1), 1972, 59-177.

[7] N. Yu, C. S. Liu and D. C. O'shea, Laser Raman spectroscopy and the conformation of insulin and proinsulin, *J Mol Biol* **70**(1), 1972, 117-132.

[8] K. Hirano, Raman spectra of DNA in aqueous solution, *Bull Chem Soc Jpn* **41**(3), 1968, 731-732.

[9] R. C. Lord and G. J. Thomas, Spectroscopic studies of molecular interaction in DNA constituents, Springer, 1968, Boston, MA,

[10] c. Lord and G. J. Thomas Jr, Raman spectral studies of nucleic acids and related molecules—I ribonucleic acid derivatives, *Spectrochim Acta, Pt A: Mol Spectrosc* **23**(9), 1967, 2551-2591.

[11] A. G. Walton, M. J. Deveney and J. L. Koenig, Raman spectroscopy of calcified tissue, *Calcif Tissue Res* **6**(1), 1970, 162-167.

[12] C. H. Liu, R. R. Alfano, W. L. Sha, H. R. Zhu, D. L. Akins, J. Cleary, et al., Human breast tissues studied by IR fourier-transform Raman spectroscopy, *Conference on lasers and electro-optics,* Optical Society of America, 1991, CWF51.

[13] J. J. Baraga, M. S. Feld and R. P. Rava, Rapid near-infrared Raman spectroscopy of human tissue with a spectrograph and CCD detector, *Appl Spectrosc* **46**(2), 1992, 187-190.

[14] A. Mizuno, H. Kitajima, K. Kawauchi, S. Muraishi and Y. Ozaki, Near-infrared fourier transform Raman spectroscopic study of human brain tissues and tumours, *J Raman Spectrosc* **25**(1), 1994, 25-29.

[15] A. Mizuno, T. Hayashi, K. Tashibu, S. Maraishi, K. Kawauchi and Y. Ozaki, Near-infrared FT-Raman spectra of the rat brain tissues, *Neurosci Lett* **141**(1), 1992, 47-52.

[16] J. Sebag, S. Nie, K. Reiser, M. A. Charles and N. Yu, Raman spectroscopy of human vitreous in proliferative diabetic retinopathy. *Invest Ophthalmol Vis Sci* **35**(7), 1994, 2976-2980.

[17] G. R. Sauer, W. B. Zunic, J. R. Durig and R. E. Wuthier, Fourier transform Raman spectroscopy of synthetic and biological calcium phosphates, *Calcif Tissue Int* **54**(5), 1994, 414-420.

[18] O. R. Scepanovic, M. Fitzmaurice, A. Miller, C. Kong, Z. I. Volynskaya, R. R. Dasari, et al., Multimodal spectroscopy detects features of vulnerable atherosclerotic plaque, *J Biomed Opt* **16**(1), 2011, 011009.

[19] R. Pandey, R. Zhou, R. Bordett, C. Hunter, K. Glunde, I. Barman, et al., Integration of diffraction phase microscopy and Raman imaging for label-free morpho-molecular assessment of live cells, *Journal of Biophotonics* **12**(4), 2019, e201800291.

[20] Y. Ozaki, Medical application of Raman spectroscopy, *Applied Spectroscopy Reviews* **24**(3-4), 1988, 259-312.

[21] P. Matousek and M. Morris, Emerging Raman applications and techniques in biomedical and pharmaceutical fields, Springer Science & Business Media, 2010, Berlin Heidelberg,

[22] Z. Movasaghi, S. Rehman and I. U. Rehman, Raman spectroscopy of biological tissues, *Applied Spectroscopy Reviews* **42**(5), 2007, 493-541.

[23] E. B. Hanlon, R. Manoharan, T. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, et al., Prospects for in vivo Raman spectroscopy, *Physics in Medicine & Biology* **45**(2), 2000, R1.

[24] A. C. S. Talari, Z. Movasaghi, S. Rehman and I. U. Rehman, Raman spectroscopy of biological tissues, *Applied Spectroscopy Reviews* **50**(1), 2015, 46-111.

[25] L. Choo-Smith, H. Edwards, H. P. Endtz, J. M. Kros, F. Heule, H. Barr, et al., Medical applications of Raman spectroscopy: From proof of principle to clinical implementation, *Biopolymers: Original Research on Biomolecules* **67**(1), 2002, 1-9.

[26] W. Xie, P. Qiu and C. Mao, Bio-imaging, detection and analysis by using nanostructures as SERS substrates, *Journal of Materials Chemistry* **21**(14), 2011, 5190-5202.

[27] W. Kiefer, Recent advances in linear and nonlinear Raman spectroscopy I, *Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering* **38**(12), 2007, 1538-1553.

[28] P. Matousek and N. Stone, Development of deep subsurface Raman spectroscopy for medical diagnosis and disease monitoring, *Chem Soc Rev* **45**(7), 2016, 1794-1802.

[29] S. K. Paidi, S. Siddhanta, R. Strouse, J. B. McGivney, C. Larkin and I. Barman, Rapid identification of biotherapeutics with label-free Raman spectroscopy, *Anal Chem* **88**(8), 2016, 4361-4368.

[30] O. Stevens, I. E. Iping Petterson, J. C. C. Day and N. Stone, Developing fibre optic Raman probes for applications in clinical spectroscopy, *Chem Soc Rev* **45**(7), 2016, 1919-1934.

[31] R. A. Crocombe, Portable spectroscopy, *Appl Spectrosc* **72**(12), 2018, 1701-1751.

[32] A. C. Croce and G. Bottiroli, Autofluorescence spectroscopy and imaging: A tool for biomedical research and diagnosis, *European Journal of Histochemistry: EJH* **58**(4), 2014,

[33] T. A. Valdez, R. Pandey, N. Spegazzini, K. Longo, C. Roehm, R. R. Dasari, et al., Multiwavelength fluorescence otoscope for video-rate chemical imaging of middle ear pathology, *Anal Chem* **86**(20), 2014, 10454-10460.

[34] R. Gautam, S. Vanga, F. Ariese and S. Umapathy, Review of multidimensional data processing approaches for Raman and infrared spectroscopy, *EPJ Techniques and Instrumentation* **2**(1), 2015, 8.

[35] S. K. Paidi, A. Rizwan, C. Zheng, M. Cheng, K. Glunde and I. Barman, Label-free Raman spectroscopy detects stromal adaptations in premetastatic lungs primed by breast cancer, *Cancer Res* **77**(2), 2017, 247-256.

[36] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan and A. Gorzsás, Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS), *Nature Protocols* **10**(2), 2015, 217.

[37] R. G. Brereton and G. R. Lloyd, Partial least squares discriminant analysis: Taking the magic away, *J Chemometrics* **28**(4), 2014, 213-225.

[38] C. Chang and C. Lin, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3), 2011, 27.

[39] J. V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, *J Clin Epidemiol* **49**(11), 1996, 1225-1231.

[40] C. Zong, M. Xu, L. Xu, T. Wei, X. Ma, X. Zheng, et al., Surface-enhanced Raman spectroscopy for bioanalysis: Reliability and challenges, *Chem Rev* **118**(10), 2018, 4946-4980.

[41] E. V. Efremov, F. Ariese and C. Gooijer, Achievements in resonance Raman spectroscopy: Review of a technique with a distinct analytical chemistry potential, *Anal Chim Acta* **606**(2), 2008, 119-134.

[42] A. Fast and E. O. Potma, Coherent Raman scattering with plasmonic antennas, *Nanophotonics* **8**(6), 2019, 991-1021.

[43] W. J. Tipping, M. Lee, A. Serrels, V. G. Brunton and A. N. Hulme, Stimulated Raman scattering microscopy: An emerging tool for drug discovery, *Chem Soc Rev* **45**(8), 2016, 2075-2089.

[44] D. Fu, F. Lu, X. Zhang, C. Freudiger, D. R. Pernik, G. Holtom, et al., Quantitative chemical imaging with multiplex stimulated Raman scattering microscopy, *J Am Chem Soc* **134**(8), 2012, 3623-3626.

[45] L. Kong, M. Ji, G. R. Holtom, D. Fu, C. W. Freudiger and X. S. Xie, Multicolor stimulated Raman scattering microscopy with a rapidly tunable optical parametric oscillator, *Opt Lett* **38**(2), 2013, 145-147.

[46] I. Barman, N. C. Dingari, A. Saha, S. McGee, L. H. Galindo, W. Liu, et al., Application of Raman spectroscopy to identify microcalcifications and

underlying breast lesions at stereotactic core needle biopsy, *Cancer Res* **73**(11), 2013, 3206-3215.

[47] C. Zheng, W. Shao, S. K. Paidi, B. Han, T. Fu, D. Wu, et al., Pursuing shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS) for concomitant detection of breast lesions and microcalcifications, *Nanoscale* **7**(40), 2015, 16960-16968.

[48] N. Stone, R. Baker, K. Rogers, A. W. Parker and P. Matousek, Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): Future possibilities for the diagnosis of breast cancer, *Analyst* **132**(9), 2007, 899-905.

[49] X. Qian, X. Peng, D. O. Ansari, Q. Yin-Goen, G. Z. Chen, D. M. Shin, et al., In vivo tumor targeting and spectroscopic detection with surface-enhanced Raman nanoparticle tags, *Nat Biotechnol* **26**(1), 2008, 83.

[50] H. Lui, J. Zhao, D. McLean and H. Zeng, Real-time Raman spectroscopy for in vivo skin cancer diagnosis, *Cancer Res* **72**(10), 2012, 2491-2500.

[51] F. M. Lyng, D. Traynor, I. R. Ramos, F. Bonnier and H. J. Byrne, Raman spectroscopy for screening and diagnosis of cervical cancer, *Analytical and Bioanalytical Chemistry* **407**(27), 2015, 8279-8289.

[52] U. Utzinger, D. L. Heintzelman, A. Mahadevan-Jansen, A. Malpica, M. Follen and R. Richards-Kortum, Near-infrared Raman spectroscopy for in vivo detection of cervical precancers, *Appl Spectrosc* **55**(8), 2001, 955-959.

[53] N. Sharma, N. Takeshita and K. Y. Ho, Raman spectroscopy for the endoscopic diagnosis of esophageal, gastric, and colonic diseases, *Clinical Endoscopy* **49**(5), 2016, 404.

[54] Y. W. Wang, S. Kang, A. Khan, P. Q. Bao and J. T. Liu, In vivo multiplexed molecular imaging of esophageal cancer via spectral endoscopy of topically applied SERS nanoparticles, *Biomedical Optics Express* **6**(10), 2015, 3714-3723.

[55] S. Duraipandian, M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, et al., Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination, *J Biomed Opt* **17**(8), 2012, 081418.

[56] K. Kong, C. Kendall, N. Stone and I. Notingher, Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection, *Adv Drug Deliv Rev* **89**2015, 121-134.

[57] A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, et al., In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy, *Cancer Res* **66**(6), 2006, 3317-3322.

[58] M. D. Keller, E. Vargis, A. Mahadevan-Jansen, N. de Matos Granja, R. H. Wilson, M. Mycek, et al., Development of a spatially offset Raman spectroscopy probe for breast tumor surgical margin evaluation, *J Biomed Opt* **16**(7), 2011, 077006.

[59] K. Kong, C. J. Rowlands, S. Varma, W. Perkins, I. H. Leach, A. A. Koloydenko, et al., Diagnosis of tumors during tissue-conserving surgery

with integrated autofluorescence and Raman scattering microscopy, *Proc Natl Acad Sci U S A* **110**(38), 2013, 15189-15194.

[60] D. W. Shipp, E. A. Rakha, A. A. Koloydenko, R. D. Macmillan, I. O. Ellis and I. Notingher, Intra-operative spectroscopic assessment of surgical margins during breast conserving surgery, *Breast Cancer Research* **20**(1), 2018, 69.

[61] R. O. Draga, M. C. Grimbergen, P. L. Vijverberg, C. F. v. Swol, T. G. Jonges, J. A. Kummer, et al., In vivo bladder cancer diagnosis by high-volume Raman spectroscopy, *Anal Chem* **82**(14), 2010, 5993-5999.

[62] M. Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, et al., Intraoperative brain cancer detection with Raman spectroscopy in humans, *Sci Transl Med* **7**(274), 2015, 274ra19.

[63] Y. Wang, S. Kang, A. Khan, G. Ruttner, S. Y. Leigh, M. Murray, et al., Quantitative molecular phenotyping with topically applied SERS nanoparticles for intraoperative guidance of breast cancer lumpectomy, *Scientific Reports* **6**2016, 21242.

[64] N. P. Reder, S. Kang, A. K. Glaser, Q. Yang, M. A. Wall, S. H. Javid, et al., Raman-encoded molecular imaging with topically applied SERS nanoparticles for intraoperative guidance of lumpectomy, *Cancer Res* **77**(16), 2017, 4506-4516.

[65] J. L. Campbell, E. D. SoRelle, O. Ilovich, O. Liba, M. L. James, Z. Qiu, et al., Multimodal assessment of SERS nanoparticle biodistribution post ingestion reveals new potential for clinical translation of Raman imaging, *Biomaterials* **135**2017, 42-52.

[66] M. Kirsch, G. Schackert, R. Salzer and C. Krafft, Raman spectroscopic imaging for in vivo detection of cerebral brain metastases, *Analytical and Bioanalytical Chemistry* **398**(4), 2010, 1707-1713.

[67] X. Bi, J. A. Sterling, A. R. Merkel, D. S. Perrien, J. S. Nyman and A. Mahadevan-Jansen, Prostate cancer metastases alter bone mineral and matrix composition independent of effects on bone architecture in mice—a quantitative study using microCT and Raman spectroscopy, *Bone* **56**(2), 2013, 454-460.

[68] C. Zhang, P. T. Winnard Jr, S. Dasari, S. L. Kominsky, M. Doucet, S. JayaRaman, et al., Label-free Raman spectroscopy provides early determination and precise localization of breast cancer-colonized bone alterations, *Chemical Science* **9**(3), 2018, 743-753.

[69] J. Horsnell, P. Stonelake, J. Christie-Brown, G. Shetty, J. Hutchings, C. Kendall, et al., Raman spectroscopy—a new method for the intra-operative assessment of axillary lymph nodes, *Analyst* **135**(12), 2010, 3042-3047.

[70] P. T. Winnard Jr, C. Zhang, F. Vesuna, J. W. Kang, J. Garry, R. R. Dasari, et al., Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes, *Oncotarget* **8**(12), 2017, 20266.

[71] X. Wang, X. Qian, J. J. Beitler, Z. G. Chen, F. R. Khuri, M. M. Lewis, et al., Detection of circulating tumor cells in human peripheral blood using surface-enhanced Raman scattering nanoparticles, *Cancer Res* **71**(5), 2011, 1526-1532.

[72] R. Pandey, S. K. Paidi, T. A. Valdez, C. Zhang, N. Spegazzini, R. R. Dasari, et al., Noninvasive monitoring of blood glucose with Raman spectroscopy, *Acc Chem Res* **50**(2), 2017, 264-272.

[73] K. Maquelin, L. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, et al., Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium, *Anal Chem* **72**(1), 2000, 12-19.

[74] S. Siddhanta, S. K. Paidi, K. Bushley, R. Prasad and I. Barman, Exploring morphological and biochemical linkages in fungal growth with Label-Free light sheet microscopy and Raman spectroscopy, *ChemPhysChem* **18**(1), 2017, 72-78.

[75] S. Kloß, B. Kampe, S. Sachse, P. Rösch, E. Straube, W. Pfister, et al., Culture independent Raman spectroscopic identification of urinary tract infection pathogens: A proof of principle study, *Anal Chem* **85**(20), 2013, 9610-9616.

[76] A. J. Hobro, A. Konishi, C. Coban and N. I. Smith, Raman spectroscopic analysis of malaria disease progression via blood and plasma samples, *Analyst* **138**(14), 2013, 3927-3933.

[77] H. P. Buschman, J. T. Motz, G. Deinum, T. J. Römer, M. Fitzmaurice, J. R. Kramer, et al., Diagnosis of human coronary atherosclerosis by morphology-based Raman spectroscopy, *Cardiovascular Pathology* **10**(2), 2001, 59-68.

[78] R. Pandey, S. K. Paidi, J. W. Kang, N. Spegazzini, R. R. Dasari, T. A. Valdez, et al., Discerning the differential molecular pathology of proliferative middle ear lesions using Raman spectroscopy, *Scientific Reports* **5**2015,

[79] K. A. Esmonde-White, G. S. Mandair, F. Raaii, J. A. Jacobson, B. S. Miller, A. G. Urquhart, et al., Raman spectroscopy of synovial fluid as a tool for diagnosing osteoarthritis, *J Biomed Opt* **14**(3), 2009, 034013.

[80] R. Ramakrishnaiah, G. u. Rehman, S. Basavarajappa, A. A. Al Khuraif, B. H. Durgesh, A. S. Khan, et al., Applications of Raman spectroscopy in dentistry: Analysis of tooth structure, *Applied Spectroscopy Reviews* **50**(4), 2015, 332-350.

[81] R. Pandey, C. Zhang, J. W. Kang, P. M. Desai, R. R. Dasari, I. Barman, et al., Differential diagnosis of otitis media with effusion using label-free Raman spectroscopy: A pilot study, *Journal of Biophotonics* **11**(6), 2018, e201700259.

[82] T. Yamamoto, T. Minamikawa, Y. Harada, Y. Yamaoka, H. Tanaka, H. Yaku, et al., Label-free evaluation of myocardial infarct in surgically excised ventricular myocardium by Raman spectroscopy, *Scientific Reports* **8**(1), 2018, 14671.

Chapter 2

Pursuing shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS) for concomitant detection of breast lesions and microcalcifications

Although tissue staining followed by morphologic identification remains the gold standard for diagnosis of most cancers, such determinations relying solely on morphology are often hampered by inter- and intra-observer variability. Vibrational spectroscopic techniques, in contrast, offer objective markers for diagnoses and can afford disease detection prior to alterations in cellular and extracellular architecture by furnishing a rapid “omics”-like view of the biochemical status of the probed specimen. Here, we report a classification approach to concomitantly detect microcalcification status and local pathological state in breast tissue, featuring a combination of vibrational spectroscopy that focuses on the tumor and its microenvironment, and multivariate data analysis of spectral markers reflecting molecular expression.

We employ the unprecedented sensitivity and exquisite molecular specificity offered by Au@SiO_2 shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS) to probe the presence of calcified deposits and

---

This chapter is a reprint of the peer-reviewed publication: Zheng C*, Shao W*, Paidi SK, Han B, Fu T, Wu D, Bi L, Xu W, Fan Z, Barman I. “Pursuing shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS) for concomitant detection of breast lesions and microcalcifications”, Nanoscale, 7, 16960-8, 2015. (* denotes equal contribution) distinguish between normal breast tissues, fibroadenoma, atypical ductal hyperplasia, ductal carcinoma in situ (DCIS), and invasive ductal carcinoma (IDC). By correlating the spectra with the corresponding histologic assessment, we developed partial least squares-discriminant analysis derived decision algorithm that provides excellent diagnostic power in the fresh frozen sections (overall accuracy of 99.4% and 93.6% using SHINs for breast lesions with and without microcalcifications, respectively). The performance of this decision algorithm is competitive with or supersedes that of analogous algorithms employing spontaneous Raman spectroscopy while enabling facile detection due to the considerably higher intensity of SHINERS. Our results pave the way for rapid tissue spectral pathology measurements using SHINERS that can offer a novel stain-free route to accurate and economical diagnoses without human interpretation.

2.1 Introduction

Examination of stained histology slides and the corresponding recognition of morphologic features remains the gold standard for definitive diagnosis and staging of most cancerous lesions. The decision-making relies extensively on a pathologist’s recognition of cell types and their morphologic patterns within the stained tissues. Although early detection and useful patient stratification impact the likelihood of cancer-free survival, manual morphological determinations involve considerable diagnostic uncertainty [1, 2] and do not elucidate the essential biochemical pathways that lead to the specific pathology. Moreover, such examination is time-consuming and expensive. While immunohistochemical approaches offer limited molecular detection, significant hurdles still remain in employing them for quantitative, automated pathology [3].

In this milieu, considerable attention has been focused on the incorporation of vibrational spectroscopic methods, as they directly offer non-perturbing molecular descriptors. Raman spectroscopy, for example, provides objective spectral markers for diagnosis of disease and may permit disease detection prior to morphologic manifestation [4]. Studies by us and others [5-14] have sought to exploit this method - due to its wealth of molecular information [15, 16] and lack of sample preparation requirements - in differentiating breast pathology and detecting the presence of microcalcifications, an important mammographic marker of breast lesions. While spontaneous Raman spectroscopy affords sub-cellular signal localization and provides an ideal in vivo characterization tool, its inherently weak signals have impeded its application for extensive tissue analysis.

Surface-enhanced Raman scattering (SERS), which was first observed in the 1970s on electrochemically roughened silver electrodes [17, 18], has alleviated this drawback with reported biomolecular detection limits beyond immunoassay sensitivities [19]. As a consequence, SERS has attracted considerable interest as an ultrasensitive and highly specific tool for non-destructive and real-time diagnosis of diseases [15, 20-28]. However, two problems still hinder the translation of SERS for biomedical applications, namely unsatisfactory substrate generality and poor measurement reproducibility [29, 30]. To overcome these drawbacks, shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS), featuring nanoparticles with Au core coated by an ultrathin shell (thickness from 2-20 nm) of silica or alumina, was proposed by Tian and co-workers at Xiamen University [31] While the gold core of the Au@SiO$_2$ shell-isolated nanoparticles (SHINs) allows obvious SERS enhancement, the shell is used to protect the gold core from degradation of the minutely fabricated nanostructures as well as preclude the interaction of the bare gold nanoparticles with the probed adsorbates. By virtue of these advantages, SHINERS has been gainfully employed in investigations of live cells [32] and, recently by us, in carcinoma studies using a ratiometric approach [33].

Using breast cancer as the paradigm, we report here an approach to concomitantly diagnose microcalcification status and local pathological state featuring SHINERS and multivariate data analysis of spectral markers reflecting molecular expression. In this article, we seek to exploit the heightened sensitivity and intrinsic specificity offered by the Au@SiO$_2$ shell-isolated nanoparticles in elucidating the biochemical content of fresh frozen specimen and, thus, in differentiating between normal breast tissue, fibroadenoma, atypical ductal hyperplasia, ductal carcinoma in situ, and invasive ductal carcinoma. By correlating the spectra with the histologic evaluation, our partial least squares - discriminant analysis derived decision algorithms show excellent diagnostic power (overall accuracy of 99.4% and 93.6% using SHINs for breast lesions with and without microcalcifications, respectively). The performance of this decision algorithm is competitive with or supersedes that of analogous algorithms developed using spontaneous Raman spectra while offering facile detection owing to the considerably higher intensity of SHINERS spectra. Notably, to the best of our knowledge, this is the first SHINERS effort that comprehensively assays microcalcifications and tissues of all key pathophysiological conditions. Taken together with the substantially higher spectral intensities for the SHINERS data in relation to that for the spontaneous Raman spectra, our findings open the door for clinical translation of SHINERS as a label-free route to accurate pathological diagnoses without human interpretation.

2.2 Materials and Methods

2.2.1 Human subjects and tissue preparation

Fresh breast tissue was collected from 72 patients who underwent surgical resection or vacuum assisted (Mammotome) biopsy at the Department of Breast Surgery, the First Hospital of Jilin University. Consent to participate in this study was obtained from every patient and the project and methodology were approved by the Ethics Committee of Jilin University. After operating, the samples were immediately frozen at -20 to -25 ^oC and two contiguous sections (6 μm thickness) were cut from each specimen using a freezing microtome (LEICA-CM3050S, Germany) in the Pathology Department, the First Hospital of Jilin University. For each tissue, one of the two adjacent sections were stained with haematoxylin and eosin (H&E) for routine histopathological analysis by three experienced breast pathologists. The other section was used for acquiring spectra from the diseased regions (identified by H&E staining of the other section) by preserving it in liquid nitrogen. Prior to analysis, the frozen section was thawed at 22 ^oC for 10 minutes and multiple spectra were collected from each tissue.

2.2.2 Instrumentation and data acquisition

A confocal Raman system (LabRAM ARAMIS, Horiba Jobin Yvon, Edison, NJ, USA) with a ~0.7 μm spatial resolution, and a 5 mW, 633nm HeNe laser as excitation source was used for the collection of Raman spectra. The detection of Raman signal was carried out with a Synapse Thermoelectric cooled charge-coupled device (CCD) camera (Horiba Jobin Yvon, Edison, NJ, USA). Raman scattered light was collected with a 50x microscope objective lens (0.50 NA, LMPLFLN, Olympus, Japan) that was also used for focusing the excitation laser light. The laser beam focused on the tissue formed a spot of 1.5 μm diameter. A 4-notch filter (Horiba Jobin Yvon, Edison, NJ, USA) was used to block strong Rayleigh-scattered light. Extended scan spectra with a spectral range of 600-1800 cm^{-1} were acquired using an integration time of 60s and 3 accumulations. The Raman shift axis was calibrated using characteristic vibration bands recorded from silicon wafer. These settings were kept constant for all the spectral measurements to enable direct comparison of spectral sensitivity and specificity, especially between spontaneous Raman and SHINERS datasets.

Locations of lesions and other possible features of interest were determined from the H&E sections and the corresponding sites on its frozen contiguous sections were probed using the Raman system. About fifteen to twenty spectra were collected from different locations for each sample to ensure representative sampling and incorporate spot-to-spot variability in the recorded signals. After obtaining spontaneous Raman spectra, SHINs were added to the surface of frozen sections and the respective SERS spectra were recorded from the same spots.

SHINs having an average Au core diameter of 55 nm and covered by 2 nm silica shells were acquired from Professor Tian's group for use in this study. On each tissue section 10 µL of 100 nM concentrated SHINs solution was added uniformly prior to collecting the spectra. **Figure 2.1** shows a representative TEM image of the SHINs used for our study.

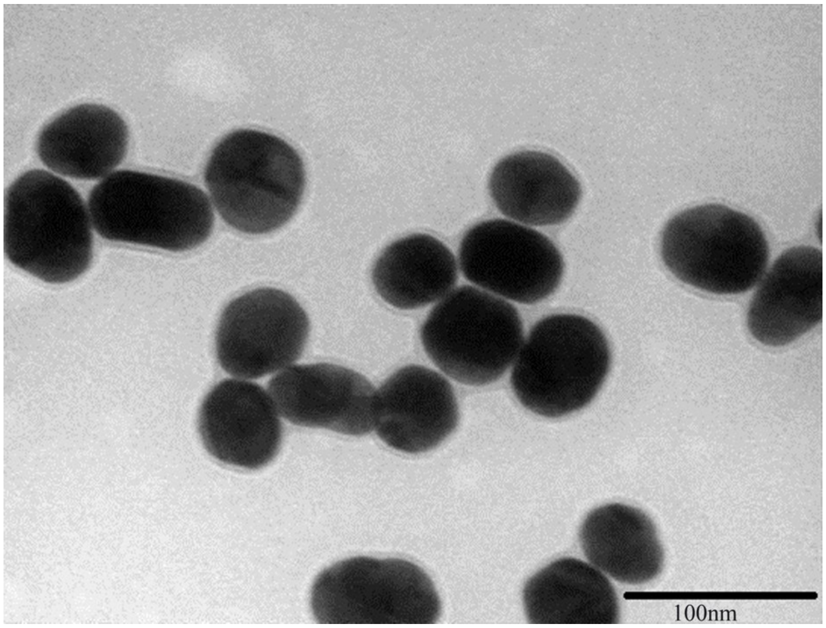


Figure 2.1: Representative TEM image of Au@SiO$_2$ shell-isolated nanoparticles (SHINs).

2.2.3 Spectral data analysis

The spectra were subjected to baseline correction by fitting and subtracting a third-order polynomial by NGSLabSpec software (Horiba Jobin Yvon, Edison, NJ, USA) and were subsequently smoothed using a 15-point averaging algorithm. These corrected spectra were used to determine the peaks characteristic of the class the spectra belong to while the intensity-normalized spectra were used in the remaining data analysis. The latter were subjected to principal component analysis (PCA), a widely used data exploration method, to capture the (subtle) spectral variances in a set of abstract orthogonal axes. For every spectrum belonging to a different pathological class, PC scores were plotted using Radviz and VizRank, radial visualization modules in Orange data mining software [34]. The nonlinear multidimensional radial visualization algorithm, RadViz, maps n data dimensions (PCs) onto two-dimensional circular space while VizRank, offers a heuristic search technique to guide the ordering of variables and evaluating the resulting radial projections by their ability to discriminate between classes. These radial visualization plots illustrate the clustering of spectra belonging to same class (assigned based on pathological assessment) and how well the classes are separated based on the PC scores. In this study, we used only the single best-ranked projection for clustering displays. Finally, the spectra belonging to various pathologies were subjected to partial least squares-discriminant analysis (PLS-DA) [35] to quantify the diagnostic power of the recorded spectra. A PLS-DA derived decision model was built based on training data and the diagnostic power was independently tested by invoking an independent set of test spectra. In particular, 60% of the spectra belonging to each class were randomly selected to train the model and the remaining 40% were tested using the model developed to check the classification capability of PLS-DA in identifying the pathology of the tissue. 1000 iterations were performed keeping the same mixture of training: test data to obtain a more robust evaluation of the classification performance.

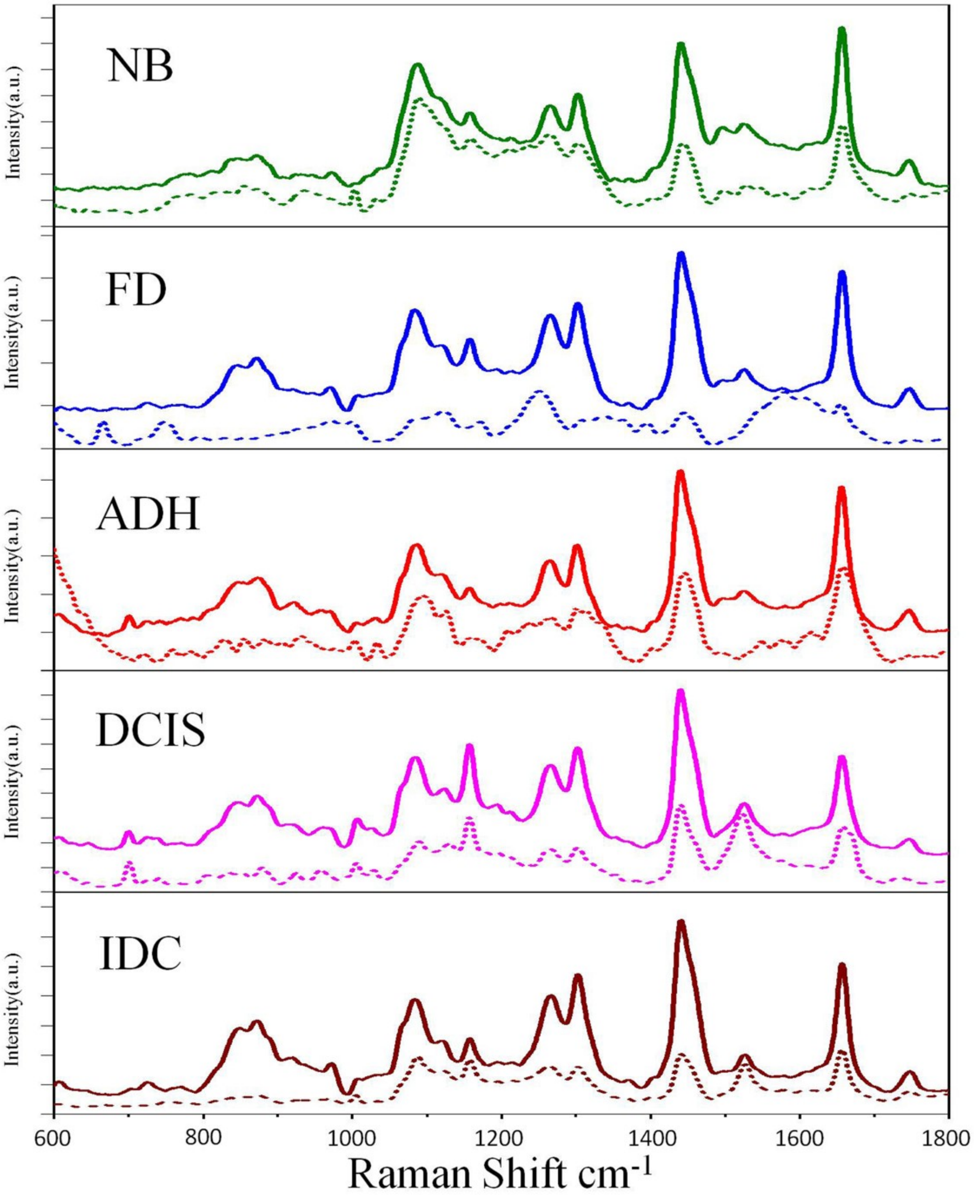


Figure 2.2: SHINERS spectra of breast cancer. Average SHINERS spectra (solid line) and spontaneous Raman spectra (dotted line) acquired from fresh frozen sections of normal breast tissue (NB) as well as those displaying the following breast lesions: fibroadenoma (FD), atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS), and invasive ductal carcinoma (IDC).

2.3 Results

The Raman spectral dataset obtained was classified as belonging to one of the following classes based on the consensus histological assessment. 17, 9, 11 and 20 tissue sections were diagnosed as fibroadenoma (FD), atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC), respectively. Additionally, 15 normal breast tissue (NB) sections were investigated for comparison with benign, premalignant and malignant pathologies. Among the diseased tissue sections, some exhibited the presence of type II microcalcifications (3 FD, 3 ADH, 5 DCIS and 4 IDC tissue sections) and were studied separately as they form an important class of pathological indicators in breast cancer diagnosis.

2.3.1 Spontaneous Raman spectroscopy and SHINERS of tissues without microcalcifications

Figure 2.2 shows the mean spontaneous Raman and SHINERS spectra of breast tissues without calcifications belonging to the different classes, i.e. NB, FD, ADH, DCIS and IDC (spectra are displayed vertically offset for visual clarity). It can clearly be seen that the Raman features were significantly enhanced by using SHINs enabling us to readily characterize the differences in spectral profiles of NB, FD, ADH, DCIS, and IDC tissue sections. The characteristic band assignments of the breast tissue, performed in our recent publication [33], indicate the presence of a similar set of chemical compounds, namely lipids, proteins, nucleic acids and β-carotenes. The main spectral features of normal breast tissue include peaks around 1300, 1442, and 1745 cm^{-1} that can be expectedly attributed to lipids (with the corresponding lack of prominent protein or nucleic acid peaks) [36, 37]. Fibroadenoma shows similar characteristics as normal breast tissue except for the Raman band at 665 cm^{-1} that is characteristic of the C-S stretching mode of cystine [36, 37]. In comparison to normal breast tissue, ADH, DCIS and IDC tissues show stronger peaks located at 1004, 1033, 1610 and 1658 cm^{-1} from vibrational modes of proteins and at 970, 1090 and 1157cm^{-1} from DNA. These are indicative of higher concentration of proteins and nucleic acids in the diseased state. Closer inspection reveals that the concentrations of proteins and nucleic acids increase gradually while lipid content reduces in the pathological transition from ADH to IDC. Additionally, DCIS and IDC tissues feature the blue shifted and broad peak of CH_2, which suggests fracture of acyl backbone of lipids and proteins [38] Also, the sensitive Raman features at 1090 cm^{-1}, characteristic of the phosphate stretching vibration in DNA, shows a perceptible shift to 1086 cm^{-1} for DCIS and IDC hinting at the possible rupture of DNA strands with the further progression of cancer. A comprehensive list of the spectral features, and the associated peak shifts, corresponding to the different pathological classes is provided in our previous report [33].

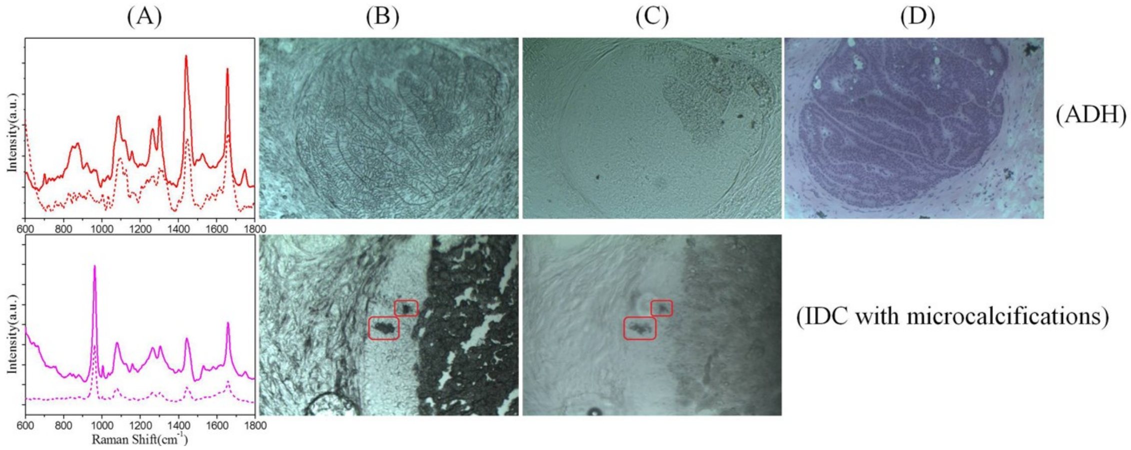


Figure 2.3: Spectroscopy histopathology comparisons for tissue section exhibiting ADH (top) and IDC with microcalcifications (bottom). (**A**) Mean SHINERS (solid line) and spontaneous Raman (dotted line) spectra acquired from the lesions; (**B**) images of the fresh frozen sections (without SHINs) that were employed in spontaneous Raman spectra acquisition. The locations of microcalcifications were highlighted by squares; (**C**) images of the fresh frozen sections after addition of SHINs. The locations of microcalcifications were highlighted by squares; (**D**) corresponding sections stained with H&E. The images represent an approximate area of 0.8 × 0.6 mm^2.

Our next aim was to correlate the spectroscopic measurements in both cases with the histological evaluation. **Figure 2.3** shows a representative set of spontaneous Raman and SHINERS spectra recorded from ADH and IDC tissue sections, respectively, with the latter showing the presence of microcalcifications. Panel (D) of the figure exhibits the corresponding H&E stained section as well as the frozen contiguous sections without the addition of SHINs (Panel (B), from which the spontaneous Raman data was recorded) and post the addition of SHINs (Panel (C), from which the SHINERS data was recorded). Based on the PCA decomposition, the most significant PC scores were then employed to create a radial visualization map to evaluate the clustering of spectra belonging to the same pathological class-or the lack thereof. **Figure 2.4** displays the cluster maps for the four types of breast lesions (FD, ADH, DCIS, IDC) using five PC scores for spontaneous Raman (A) and SHINERS spectra (B), respectively. (The corresponding PC loadings are given in **Fig. S2.1** of Appendix.)

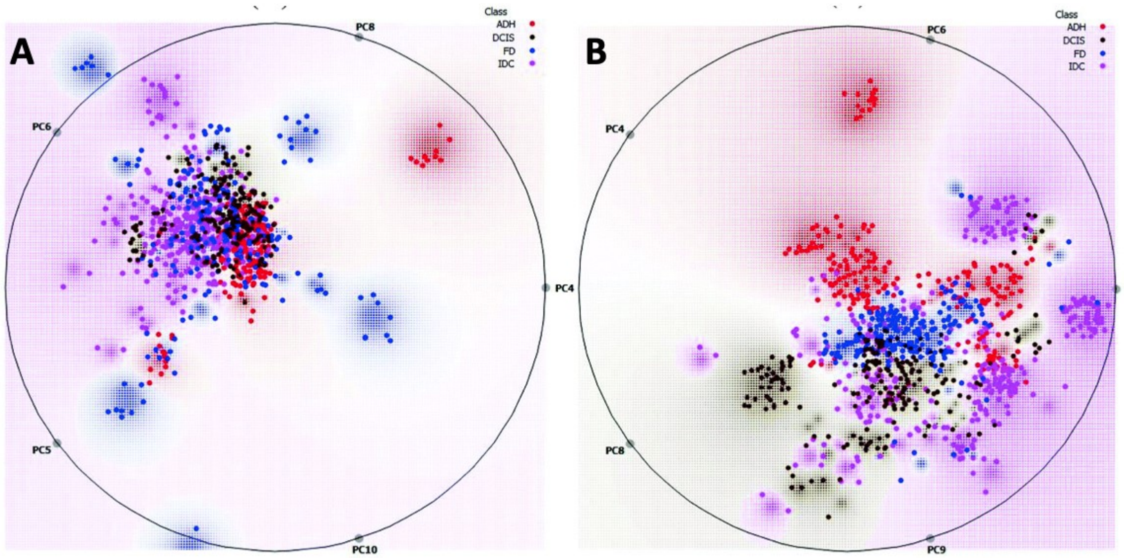


Figure 2.4: Multi-dimensional radial visualization plots for principal component (PC) scores obtained from SHINERS spectra of breast cancer. The specific PCs were selected using a heuristic search algorithm that enables maximal separation of the four diseased groups. (**A**) Spontaneous Raman spectra; (**B**) SHINERS spectra. The plots highlight the degree of clustering of sites belonging to a specific pathological class in the absence of microcalcifications for the two spectral datasets.

From the figure, we observe that the sites belonging to a particular class show a tendency to cluster together, although there is considerable overlap between few of the classes particularly for (A) (spontaneous Raman spectra-based PC scores map). The Raman spectra-derived overlap is consistent with our previous report [12] where significant segmentation between the pathologies was not observed visually-despite the absence of ADH from that dataset. This overlap is significantly reduced in **Fig. 2.4B**, which yields reasonably satisfactory (though not perfect) class boundaries. We attribute this qualitative enhancement in spectra-based class separation to the enhancement of the subtle vibrational mode-specific spectral features by SHINs.

Table 2.1. Summary of classification accuracies for PLS-DA derived decision algorithms featuring spontaneous Raman and SHINERS spectra, respectively. Standard deviations are noted in the parentheses. In this analysis, spectra and identity labels corresponding to normal breast (NB), fibroadenoma (FD), acute ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC) were included.

	Accuracy of spontaneous Raman spectroscopy	**Accuracy of SHINERS**
NB	72.8% (5.7%)	82.0% (5.1%)
FD	91.0% (4.4%)	92.3% (2.5%)
ADH	97.2% (1.7%)	97.5% (1.6%)
DCIS	87.4% (6.0%)	92.7% (3.9%)
IDC	96.9% (2.0%)	91.8% (4.0%)

Quantitative assessment of the diagnostic capability can be obtained from the results of PLS-DA-derived decision algorithm summarized in **Table 2.1**. Using independent training and test sets, we computed the overall accuracy of the PLS-DA derived decision algorithms in classifying NB, FD, ADH, DCIS and IDC to be 72.8%, 91.0%, 97.2%, 87.4% and 96.9%, respectively using spontaneous Raman spectroscopy. The analogous PLS-DA derived decision models for the SHINERS data yielded mean classification accuracy of 82.0%, 92.3%, 97.5%, 92.7% and 91.8%, respectively. Despite the extremely challenging cohort studied here (as reflected by the presence of ADH, DCIS and IDC sites), Raman spectroscopy (both spontaneous and enhanced) offer significant spectral markers of the (downstream) molecular determinants in the epithelial and stromal tissue components. Specifically, when used in conjunction with chemometric methods, these spectral profiles enable objective disease detection with high diagnostic accuracy. The overall classification performance rose for all categories from spontaneous Raman to SHINERS, barring IDC, with significant improvements for normal breast tissue and DCIS segmentation. The inferior performance of DCIS classification using spontaneous Raman spectra is reflective of the large spread observed in **Fig. 2.4A**, while the reduced classification accuracy of IDC for SHINERS can also be predicted based on the dispersion of the IDC sites in **Fig. 2.4B**.

2.3.2 Spontaneous Raman spectroscopy and SHINERS of tissues with type II microcalcifications

Raman spectroscopy has been previously investigated in formalin-fixed, paraffin-embedded sections for detection of type I and II breast microcalcifications as well as for the classification of type II microcalcifications related to benign and malignant lesions [39]. More recently, we have also developed Raman spectroscopy algorithms to detect microcalcifications in fresh breast needle biopsy tissue cores [11].

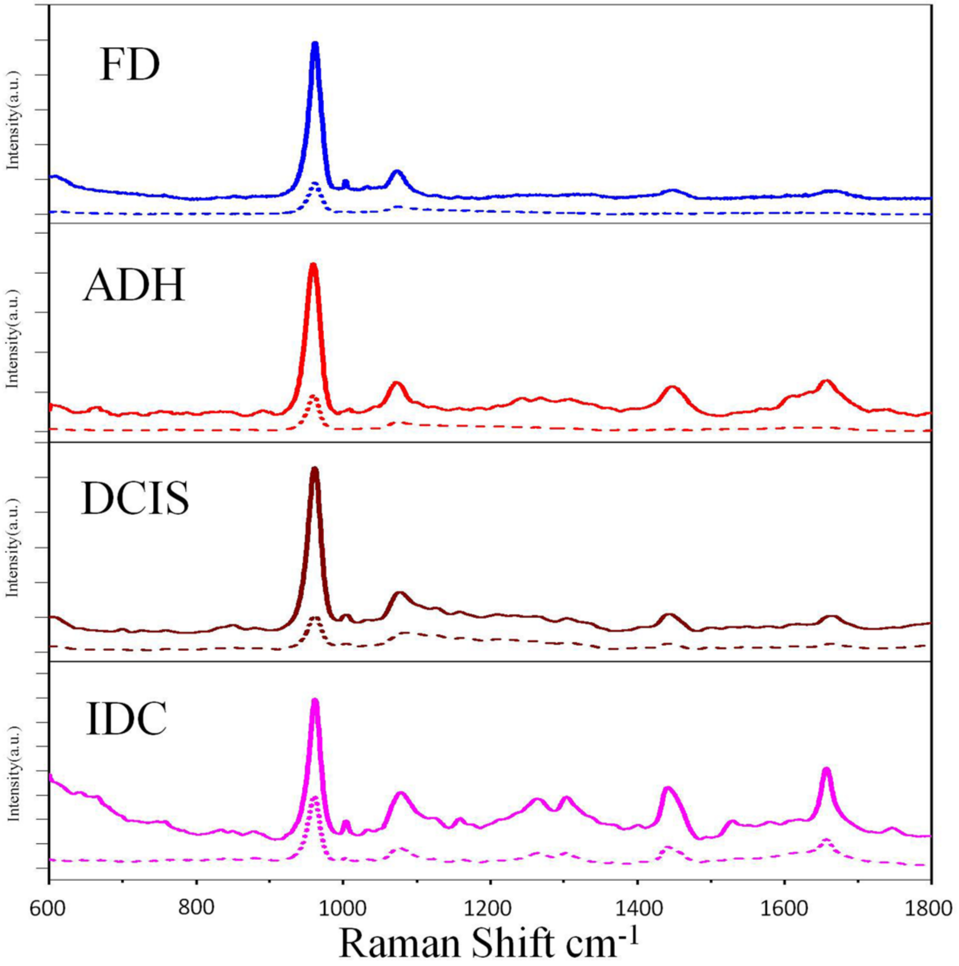


Figure 2.5: SHINERS spectra of breast microcalcifications. Average SHINERS spectra (solid line) and spontaneous Raman spectra (dotted line) acquired from tissue sites displaying type II microcalcifications with the following breast pathologies: fibroadenoma (FD), atypical ductal hyperplasia (ADH), ductal carcinoma *in situ* (DCIS), and invasive ductal carcinoma (IDC).

Here, **Fig. 2.5** shows the mean spectra of tissues showing presence of microcalcifications with and without SHINs measured from fresh frozen sections of FD, ADH, DCIS, and IDC tissue. The major peak at *ca.* 960 cm^{-1} is a well-known calcium hydroxyapatite peak and can be ascribed to the totally symmetric stretching mode of the “free” tetrahedral phosphate ion [39-41]. The other noticeable peaks appear at 1002, 1072, 1156, 1446, and 1663 cm^{-1} that are attributed to amino acid residue (phenylalanine) [42], nucleic acids, lipids, carotenoids, and amide-I, respectively [36, 37]. FD and ADH microcalcifications spectra exhibit a peak at 1072 cm^{-1}, which corresponds to the O-P-O stretch of nucleic acids, whereas in DCIS and IDC tissues the feature is observed to be red-shifted to 1078 cm^{-1}. The spectra belonging to ADH show a stronger characteristic amide-I peak at 1657 cm^{-1} while DCIS tissue sites harboring microcalcifications show the presence of relatively stronger peaks corresponding to amino acid residue (phenylalanine) at 1003, 1031 cm^{-1} and lipids (CH_2 and CH_3 bending) at 1301, 1441 cm^{-1}. Also, compared to the DCIS sites, the spectra recorded from IDC sites with microcalcifications show stronger nucleic acids peaks and weaker lipids peaks.

Figure 2.6 shows the corresponding radial visualization plot employing selected PC scores to depict maximum class separation of the four diseased classes. (The corresponding PC loadings are given in **Fig. S2.1** of Appendix.) Once again, it is evident that the addition of SHINs improves the separation between sites belonging to the same pathological class. A comparison of **Fig. 2.4** and **Fig. 2.6** reveals that the separation for both spontaneous Raman and SHINERS data is much improved for the tissue sites harboring microcalcifications. To quantitatively evaluate the diagnostic advantage provided by microcalcifications in classifying the diseased tissues, PLS-DA was employed to develop decision models, as detailed earlier. The results obtained are summarized in **Table 2.2**. In this analysis restricted to sites harboring microcalcifications alone, classification accuracies of 99.0%, 98.1%, 96.9% and 100% were achieved for the classes FD, ADH, DCIS and IDC respectively, using spontaneous Raman spectroscopy. The PLS-DA derived decision models for SHINERS, on the other hand, yielded accuracies of 99.9%, 99.9%, 97.5% and 98.2% for the same pathologies. Clearly, the multivariate analysis corroborates the visual findings that the accuracy of classification is higher for tissues harboring microcalcifications in comparison with those that show no microcalcifications. Detection of such biomineralization processes is of value not only in examining breast pathology but also in recognizing mineralized deposits in other organs, including the thyroid [43] and middle ear [44].

2.4 Discussion

The results of the present study demonstrate the ability of label-free Raman spectroscopy to differentiate closely related breast pathologies. Furthermore, the improvement of the classification results and the class separation in radial plots with the addition of SHINs underscores their diagnostic utility in tissue spectropathology. These two label-free approaches collectively offer a powerful tool to measure both localized cellular changes (especially in the case of SHINERS, since SERS presents a very surface selective method due to the decay of field enhancement with r^{-3}) [45] and the tissue microenvironment. While a critical role for the stromal environment is acknowledged in the onset and progression of cancer [46], existing diagnostic methods largely focus on the characterization of epithelial cells-a barrier that can be overcome by the employment of the employed photonic tools.

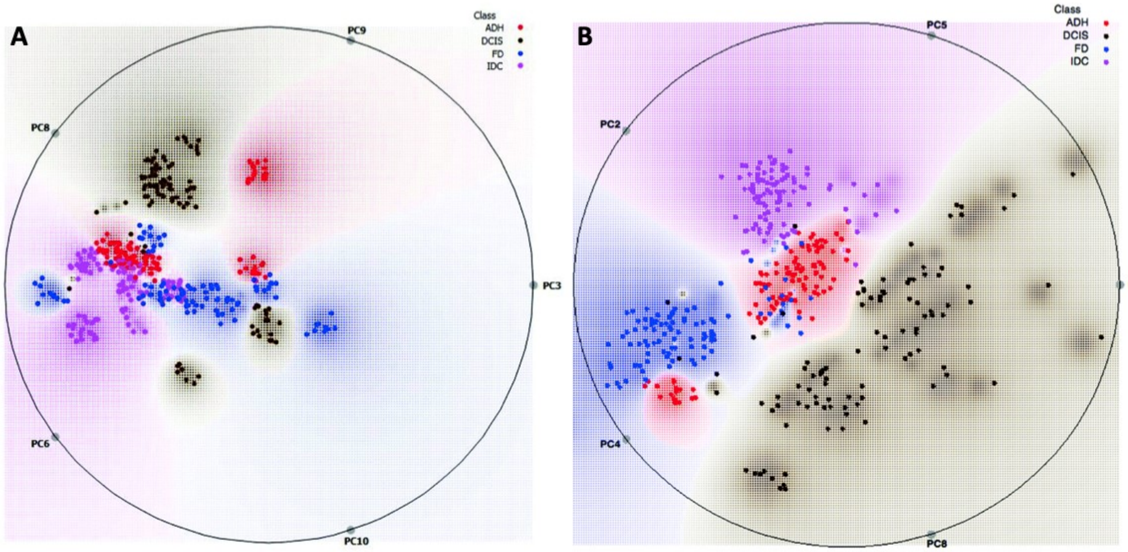


Figure 2.6: Multi-dimensional radial visualization plots for principal component (PC) scores obtained from SHINERS spectra of breast microcalcifications. Radial visualization map using principal component (PC) scores that provide maximal separation between tissues sites belonging to different pathologies and harboring microcalcifications. (**A**) Spontaneous Raman spectra; (**B**) SHINERS spectra.

Table 2.2: Summary of classification accuracies for PLS-DA derived decision algorithms featuring spontaneous Raman and SHINERS spectra, respectively. Standard deviations are noted in the parentheses. In this analysis, spectra and identity labels corresponding to tissue sites harboring microcalcifications and displaying the following pathologies were included: fibroadenoma (FD), acute ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC).

	Accuracy of spontaneous Raman spectroscopy	**Accuracy of SHINERS**
FD	99.0% (1.5%)	99.9% (0.2%)
ADH	98.1% (6.3%)	99.9% (1.7%)
DCIS	96.9% (2.6%)	97.5% (2.8%)
IDC	100.0% (0%)	98.2% (1.8%)

Significantly, the developed multivariate decision models were able to exploit the molecular structural differences - as indicated by spectral peak shifts, band shapes and relative intensity changes occurring across the Raman spectrum - to identify ADH, a benign lesion of the breast that indicates an increased risk of breast cancer. This differentiation has represented an outstanding challenge for most non-invasive photonic tools, due to its similarity to low-grade DCIS - cytologically, architecturally and on a molecular basis. In fact, early and accurate detection of ADH and DCIS prior to their progression into invasive disease represents a key goal in breast cancer diagnostics [47]. We observed the spectral markers representative of vibrational modes of proteins and nucleic acids increase gradually while those of lipids reduce in the pathological progression from ADH to IDC. In addition, the CH_2 and the C=O stretching mode of phospholipids showed a consistent 2-3 cm^{-1} blue shift in DCIS and IDC tissues, due to lipid peroxidation that occurs in the process of canceration [38]. Based on these findings, this study provides a significant extension to the literature by accurately differentiating these closely related pathologies (especially ADH and DCIS) in fresh frozen tissues (rather than in deparaffinized sections that may introduce spectral artifacts [48]. Perhaps surprisingly, the accuracy of classification for normal breast tissue using both spontaneous Raman spectroscopy and SHINERS was found to be lower than diseased tissues (ca. 73% and 82% respectively) in the recorded dataset. We attribute this to the probable presence of some uninvolved tissue sites in the diseased pathology categories that would then skew the classification model trained on spectra originating from such sites.

We have also discovered the presence of subtle, but reproducible, differences in the chemical composition of tissue sites harboring type II microcalcifications among the four kinds of studied lesions. It was observed that the peak of O-P-O stretch of nucleic acids at 1072 cm^{-1} in the spectra of FD and ADH tissues shifted to 1078 cm^{-1} for DCIS and IDC tissues, as a result of the changed calcification compositions of different kinds of diseases. Moreover, the ADH sites exhibit a significant amide-I peak at 1657 cm^{-1}; in contrast, the same peak loses its intensity in DCIS and IDC tissues owing to the process of canceration whereby the protein conformation is damaged to varying degrees.

A further important goal of this work was to compare the diagnostic power of SHINERS based decision models to that derived from spontaneous Raman data. We find that, on average, the performance of the SHINERS decision models supersedes that of its spontaneous Raman counterpart. This is also reflected in the clearer clustering of the SHINERS based PC scores corresponding to the different pathologies in the radial visualization plots. A slight decrease in the classification accuracy of IDC sites (with and without microcalcifications) is observed for SHINERS in relation to spontaneous Raman measurements. While the precise rationale for this is the subject of ongoing investigations in a larger patient cohort, a plausible explanation lies in the (sharper) differences in the nature of molecular changes that appear in IDC among different patients due to its higher degree of malignancy. Following the enhancement of spectral features by SHINERS, these inter-lesion differences are amplified resulting in worse performance of the multivariate classification models.

Finally, the intensity ratios of the key lipid, protein and nucleic acid peaks underscore the improvement in signal sensitivity due to the incorporation of SHINs and indicate that a significant reduction in SHINERS acquisition time could still result in nearly equivalent spectral profiles as obtained in the spontaneous Raman measurements. This has substantive implications for Raman spectroscopic imaging, which would overcome the sampling limitations of single point recordings and combine the molecular and spatial information content into a hyperspectral dataset. Unlike targeted SERS probes, where the SERS reporters transduce the presence of the biochemical analytes of interest into measurable signals, SHINERS allows the enhancement of the intrinsic vibrational modes of all the molecular species present in the probed specimen. This ensures the retention of the exquisite molecular specificity of spontaneous Raman spectroscopy thereby offering a unique combination of signal enhancement with facile readout and extensive multiplexing.

2.5 Appendix

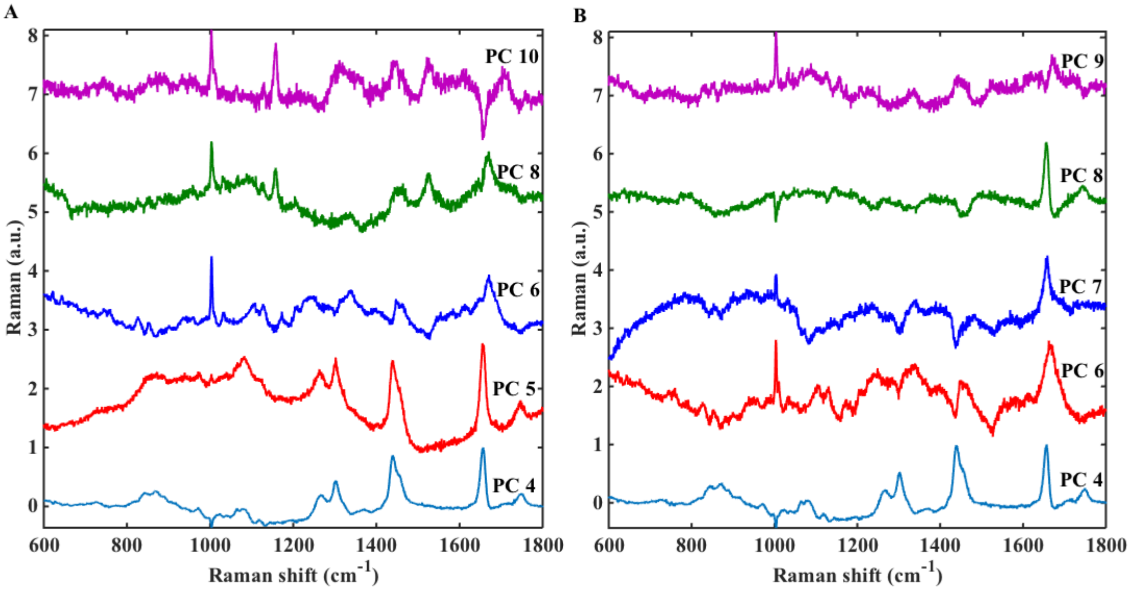


Figure S2.1: The principal component loadings (PCs) used to draw radial visualization plots used in Fig. 2.4. (**A**) Spontaneous Raman spectra from tissues without microcalcifications; (**B**) SHINERS spectra from tissues without microcalcifications.

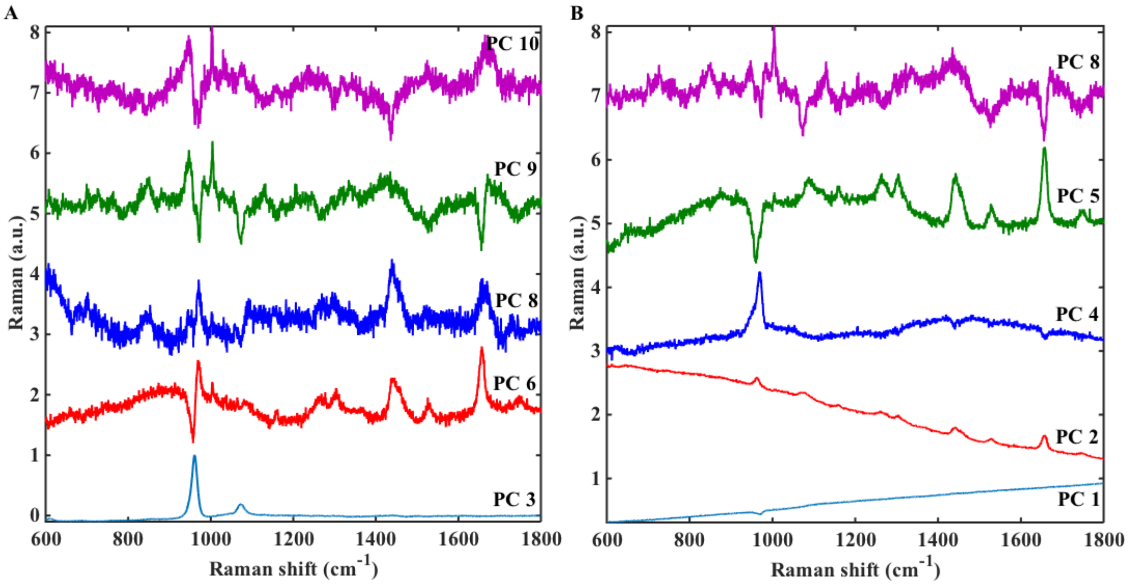


Figure S2.2: The principal component loadings (PCs) used to draw radial visualization plots used in Fig. 2.6. (**A**) Spontaneous Raman Spectra from tissues exhibiting microcalcifications; (**B**) SHINERS spectra from tissues exhibiting microcalcifications.

Bibliography

[1] W. C. Allsbrook, K. A. Mangold, M. H. Johnson, R. B. Lane, C. G. Lane, M. B. Amin, et al., Interobserver reproducibility of gleason grading of prostatic carcinoma: Urologic pathologists, *Hum Pathol* **32**(1), 2001, 74-80.

[2] M. R. Emmert-Buck, R. L. Strausberg, D. B. Krizman, M. F. Bonaldo, R. F. Bonner, D. G. Bostwick, et al., Molecular profiling of clinical tissue specimens - feasibility and applications, *Am J Pathol* **156**(4), 2000, 1109-1115.

[3] D. C. Fernandez, R. Bhargava, S. M. Hewitt and I. W. Levin, Infrared spectroscopic imaging for histopathologic recognition, *Nat Biotechnol* **23**(4), 2005, 469-474.

[4] C. A. Lieber, H. E. Nethercott and M. H. Kabeer, Cancer field effects in normal tissues revealed by raman spectroscopy, *Biomed Opt Express* **1**(3), 2010, 975-982.

[5] S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari and M. S. Feld, Diagnosing breast cancer by using raman spectroscopy, *Proceedings of the National Academy of Sciences of the United States of America* **102**(35), 2005, 12371-12376.

[6] P. Matousek and N. Stone, Recent advances in the development of raman spectroscopy for deep non-invasive medical diagnosis, *Journal of biophotonics* **6**(1), 2013, 7-19.

[7] D. Evers, B. Hendriks, G. Lucassen and T. Ruers, Optical spectroscopy: Current advances and future applications in cancer diagnostics and therapy, *Future oncology (London, England)* **8**(3), 2012, 307-320.

[8] C. M. Krishna, J. Kurien, S. Mathew, L. Rao, K. Maheedhar, K. K. Kumar, et al., Raman spectroscopy of breast tissues, *Expert Rev Mol Diagn* **8**(2), 2008, 149-166.

[9] C. Hu, J. Wang, C. Zheng, S. Xu, H. Zhang, Y. Liang, et al., Raman spectra exploring breast tissues: Comparison of principal component analysis and support vector machine-recursive feature elimination, *Medical physics* **40**(6), 2013, 063501.

[10] Barman, N. C. Dingari, A. Saha, S. Mcgee, L. H. Galindo, W. Liu, et al., Application of raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy, *Cancer Res* **73**(11), 2013, 3206-3215.

[11] Saha, I. Barman, N. C. Dingari, S. Mcgee, Z. Volynskaya, L. H. Galindo, et al., Raman spectroscopy: A real-time tool for identifying microcalcifications during stereotactic breast core needle biopsies, *Biomed Opt Express* **2**(10), 2011, 2792-2803.

[12] N. C. Dingari, I. Barman, A. Saha, S. Mcgee, L. H. Galindo, W. Liu, et al., Development and comparative assessment of raman spectroscopic classification algorithms for lesion discrimination in stereotactic breast biopsies with microcalcifications, *Journal of biophotonics* **6**(4), 2013, 371-381.

[13] Saha, I. Barman, N. C. Dingari, L. H. Galindo, A. Sattar, W. Liu, et al., Precision of raman spectroscopy measurements in detection of

microcalcifications in breast needle biopsies, *Anal Chem* **84**(15), 2012, 6715-6722.

[14] R. Sathyavathi, A. Saha, J. S. Soares, N. Spegazzini, S. Mcgee, R. Rao Dasari, et al., Raman spectroscopic sensing of carbonate intercalation in breast microcalcifications at stereotactic biopsy, *Scientific Reports* **5**(2015, 9907.

[15] H. Abramczyk and B. Brozek-Pluska, Raman imaging in biochemical and biomedical applications. Diagnosis and treatment of breast cancer, *Chemical Reviews* **113**(8), 2013, 5766-5781.

[16] S. Devpura, K. N. Barton, S. L. Brown, O. Palyvoda, S. Kalkanis, V. M. Naik, et al., Vision 20/20: The role of raman spectroscopy in early stage cancer detection and feasibility for application in radiation therapy response assessment, *Medical physics* **41**(5), 2014, 050901.

[17] M. Fleischmann, P. J. Hendra and A. J. Mcquillan, Raman spectra of pyridine adsorbed at a silver electrode, *Chemical Physics Letters* **26**(2), 1974, 163-166.

[18] D. L. Jeanmaire and R. P. Van Duyne, Surface raman spectroelectrochemistry: Part i. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode, *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* **84**(1), 1977, 1-20.

[19] D. S. Grubisha, R. J. Lipert, H. Y. Park, J. Driskell and M. D. Porter, Femtomolar detection of prostate-specific antigen: An immunoassay based

on surface-enhanced raman scattering and immunogold labels, *Anal Chem* **75**(21), 2003, 5936-5943.

[20] M. Y. Sha, H. Xu, S. G. Penn and R. Cromer, Sers nanoparticles: A new optical detection modality for cancer diagnosis, *Nanomedicine (London, England)* **2**(5), 2007, 725-734.

[21] Q. Tu and C. Chang, Diagnostic applications of raman spectroscopy, *Nanomedicine (London, England)* **8**(5), 2012, 545-558.

[22] L. Dykman and N. Khlebtsov, Gold nanoparticles in biomedical applications: Recent advances and perspectives, *Chemical Society reviews* **41**(6), 2012, 2256-2282.

[23] H. Chon, S. Lee, S. Y. Yoon, S. I. Chang, D. W. Lim and J. Choo, Simultaneous immunoassay for the detection of two lung cancer markers using functionalized sers nanoprobes, *Chemical communications (Cambridge, England)* **47**(46), 2011, 12515-12517.

[24] Lin, R. Chen, S. Feng, J. Pan, B. Li, G. Chen, et al., Surface-enhanced raman scattering spectroscopy for potential noninvasive nasopharyngeal cancer detection, *Journal of Raman Spectroscopy* **43**(4), 2012, 497-502.

[25] Yang, Z. Wang, S. Zong, C. Song, R. Zhang and Y. Cui, Distinguishing breast cancer cells using surface-enhanced raman scattering, *Analytical and bioanalytical chemistry* **402**(3), 2012, 1093-1100.

[26] P. Wu, Y. Gao, H. Zhang and C. Cai, Aptamer-guided silver-gold bimetallic nanostructures with highly active surface-enhanced raman scattering for

specific detection and near-infrared photothermal therapy of human breast cancer cells, *Anal Chem* **84**(18), 2012, 7692-7699.

[27] Y. Yang, F. Li, L. Gao, Z. Wang, M. J. Thrall, S. S. Shen, et al., Differential diagnosis of breast cancer using quantitative, label-free and molecular vibrational imaging, *Biomed Opt Express* **2**(8), 2011, 2160-2174.

[28] Li, J. W. Kang, S. Sukumar, R. R. Dasari and I. Barman, Multiplexed detection of serological cancer markers with plasmon-enhanced raman spectro-immunoassay, *Chemical Science* **6**(7), 2015, 3906-3914.

[29] J. F. Li, X. D. Tian, S. B. Li, J. R. Anema, Z. L. Yang, Y. Ding, et al., Surface analysis using shell-isolated nanoparticle-enhanced raman spectroscopy, *Nature protocols* **8**(1), 2013, 52-65.

[30] S. Park, P. X. Yang, P. Corredor and M. J. Weaver, Transition metal-coated nanoparticle films: Vibrational characterization with surface-enhanced raman scattering, *J Am Chem Soc* **124**(11), 2002, 2428-2429.

[31] J. F. Li, Y. F. Huang, Y. Ding, Z. L. Yang, S. B. Li, X. S. Zhou, et al., Shell-isolated nanoparticle-enhanced raman spectroscopy, *Nature* **464**(7287), 2010, 392-395.

[32] J. F. Li, X. D. Tian, S. B. Li, J. R. Anema, Z. L. Yang, Y. Ding, et al., Surface analysis using shell-isolated nanoparticle-enhanced raman spectroscopy, *Nature protocols* **8**(1), 2013, 52-65.

[33] C. Zheng, L. Liang, S. Xu, H. Zhang, C. Hu, L. Bi, et al., The use of au@sio$_2$ shell-isolated nanoparticle-enhanced raman spectroscopy for human breast

cancer detection, *Analytical and bioanalytical chemistry* **406**(22), 2014, 5425-5432.

[34] J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, et al., Orange: Data mining toolbox in python, *J Mach Learn Res* **14**(2013, 2349-2353.

[35] R. G. Brereton and G. R. Lloyd, Partial least squares discriminant analysis: Taking the magic away, *J Chemometr* **28**(4), 2014, 213-225.

[36] R. E. Kast, G. K. Serhatkulu, A. Cao, A. K. Pandya, H. Dai, J. S. Thakur, et al., Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model, *Biopolymers* **89**(3), 2008, 235-241.

[37] Stone, C. Kendall, J. Smith, P. Crow and H. Barr, Raman spectroscopy for identification of epithelial cancers, *Faraday discussions* **126**(2004, 141-157; discussion 169-183.

[38] H. Abramczyk, I. Placek, B. Brozek-Pluska, K. Kurczewski, Z. Morawiecc and M. Tazbir, Human breast tissue cancer diagnosis by raman spectroscopy, *Spectr-Int J* **22**(2-3), 2008, 113-121.

[39] S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari and M. S. Feld, Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using raman spectroscopy, *Cancer Res* **62**(18), 2002, 5375-5380.

[40] L. Liang, C. Zheng, H. Zhang, S. Xu, Z. Zhang, C. Hu, et al., Exploring type ii microcalcifications in benign and premalignant breast lesions by shell-

isolated nanoparticle-enhanced raman spectroscopy (shiners), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **132**(0), 2014, 397-402.

[41] R. Baker, P. Matousek, K. L. Ronayne, A. W. Parker, K. Rogers and N. Stone, Depth profiling of calcifications in breast tissue using picosecond kerr-gated raman spectroscopy, *The Analyst* **132**(1), 2007, 48-53.

[42] Marcelo, R. Leandro, A. Emília Ângelo Loschiavo, S. Ana Maria Do Espírito, S. Edson Aparecido Pereira Dos, B. Renata Andrade, et al., Raman spectroscopy study of breast disease, *Theoretical Chemistry Accounts* **125**(3-6), 2009, 329-334.

[43] M. L. C. Khoo, S. L. Asa, I. J. Witterick and J. L. Freeman, Thyroid calcification and its association with thyroid carcinoma, *Head & Neck* **24**(7), 2002, 651-655.

[44] R. Pandey, S. K. Paidi, J. W. Kang, N. Spegazzini, R. R. Dasari, T. A. Valdez, et al., Discerning the differential molecular pathology of proliferative middle ear lesions using raman spectroscopy, *Scientific Reports* **5**(2015, 13305.

[45] L. Stiles, J. A. Dieringer, N. C. Shah and R. P. Van Duyne, Surface-enhanced raman spectroscopy, *Annual Review of Analytical Chemistry* **1**(1), 2008, 601-626.

[46] H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. Van De Vijver, et al., Systematic analysis of breast cancer morphology uncovers stromal features associated with survival, *Science Translational Medicine* **3**(108), 2011, 108ra113-108ra113.

[47] Bombonati and D. C. Sgroi, The molecular pathology of breast cancer progression, *The Journal of pathology* **223**(2), 2011, 307-317.

[48] R. Wiens, M. Rak, N. Cox, S. Abraham, B. H. J. Juurlink, W. M. Kulyk, et al., Synchrotron ftir microspectroscopic analysis of the effects of anti-inflammatory therapeutics on wound healing in laminectomized rats, *Analytical and bioanalytical chemistry* **387**(5), 2007, 1679-1689.

Chapter 3

Mapping the genetic basis of breast microcalcifications and their role in metastasis

Breast cancer screening and early stage diagnosis is typically performed by X-ray mammography, which detects microcalcifications. Despite being one of the most reliable features of nonpalpable breast cancer, the processes by which these microcalcifications form are understudied and largely unknown. In the current work, we have investigated the genetic drivers for the formation of microcalcifications in breast cancer cell lines and have investigated their involvement in disease progression. We have shown that stable silencing of the Osteopontin (OPN) gene decreased the formation of hydroxyapatite in MDA-MB-231 breast cancer cells in response to osteogenic cocktail. In addition, OPN silencing reduced breast cancer cell migration. Furthermore, breast cancer cells that had spontaneously metastasized to the lungs in a mouse model of breast cancer had largely elevated OPN levels, while circulating tumor cells in the same mouse model contained intermediately increased OPN levels as compared to parental cells. The observed dual roles of the OPN gene reveal the existence of a direct relationship between calcium deposition and the ability of breast cancer cells to metastasize to distant organs, mediated by common genetic factors.

This chapter is a reprint of the peer-reviewed publication: Rizwan A*, Paidi SK*, Zheng C*, Cheng M, Fan Z, Barman I, Glunde K. “Mapping the genetic basis of breast microcalcifications and their role in metastasis”, Scientific Reports, 8:11067, 2018. (* denotes equal contribution)

3.1 Introduction

Breast cancer is the most common malignancy in women with an incidence rate of about 120 in 100,000 women in the United States [1]. The 5-year survival rate of breast cancer patients drops from ~99% for Stage I patients, to ~27% for Stage IV disease, and thus necessitates early detection [1]. Mammography to reveal microcalcifications in the breasts is the most widely used tool in breast cancer screening and for the initial diagnosis of non-palpable breast tumors [2]. The use of microcalcifications as a reliable biomarker of breast cancer has also been questioned due to their association with both benign and malignant lesions, which leads to unnecessary biopsies [3,4]. Specifically, microcalcifications that are composed of calcium hydroxyapatite are found in both benign breast lesions and breast cancers whereas those constituted by calcium oxalate crystals are largely indicative of benign lesions. For several decades, research has mostly focused on recognizing the various morphologies that microcalcifications can have in breast tissue and their correlation with the degree of malignancy [5]. Emerging evidence from us and others suggests that higher hydroxyapatite content in mammary microcalcifications is a marker for malignant disease whereas lower hydroxyapatite and a relatively higher calcium carbonate content is characteristic of benign breast lesions [6]. Yet, such studies have provided limited information about the mechanisms governing the genesis of microcalcifications and their role in disease progression.

After having collectively been viewed as a result of cellular degeneration, a paradigm shift has recently been proposed that specific type(s) of microcalcifications are products of active cellular processes and may result from processes similar to those involved in physiological bone mineralization [7,8]. Bellahacene et al. reported increased expression of bone matrix proteins, which are typically involved in physiological bone mineralization, in human breast cancer cells, and speculated that they may have a role in hydroxyapatite formation [9,10]. Recently, Scimeca et al. showed that, under specific stimuli, epithelial cells undergoing epithelial-mesenchymal-transition (EMT) transform themselves into cells with an osteoblast-like phenotype, and are able to contribute to the production of breast microcalcifications [11]. They further demonstrated that the localization of hydroxyapatite in these cancer cells was similar to that in osteoblasts. These observations suggest that to understand the role of microcalcifications in breast cancer, it is imperative to systematically explore the genetic basis of their formation, subsequent transportation into the extracellular matrix and involvement in metastatic cancer progression.

In the current study, we seek to identify and study key genetic factors that guide the formation of microcalcifications from mammary cells, and their relationship with the migration capabilities of human breast cancer cells. To achieve this, we have examined publicly available microarray data sets for potential gene candidates in a blinded and unbiased fashion, which are differentially expressed in aggressive human breast cancer cell lines that typically develop microcalcifications in vitro compared to non-aggressive lines. The obtained list of candidate genes was further refined by selecting genes encoding proteins that have putative roles in tissue or cellular microcalcification. We identified the SPP1 gene encoding osteopontin (OPN) to be the most differentially expressed gene characteristic of aggressive cell lines in our list of genes. Osteopontin (OPN) is a secreted soluble glycoprotein that is present in most body fluids including milk and serum [12]. It is overexpressed in a number of different carcinomas and has previously been implicated as an enhancer of mineralization in human breast cancer samples9. Secreted OPN interacts with multiple cell surface receptors, including various integrins (integrin β1, integrin β3) and CD44 [13]. Several studies have proposed a link between OPN and cancer [14–20]. This link, in particular to metastasis, is based on the binding of OPN to cell surface receptors such as CD44, which is critical to EMT initiation and cell-matrix adhesion in various types of primary tumors [21–23].

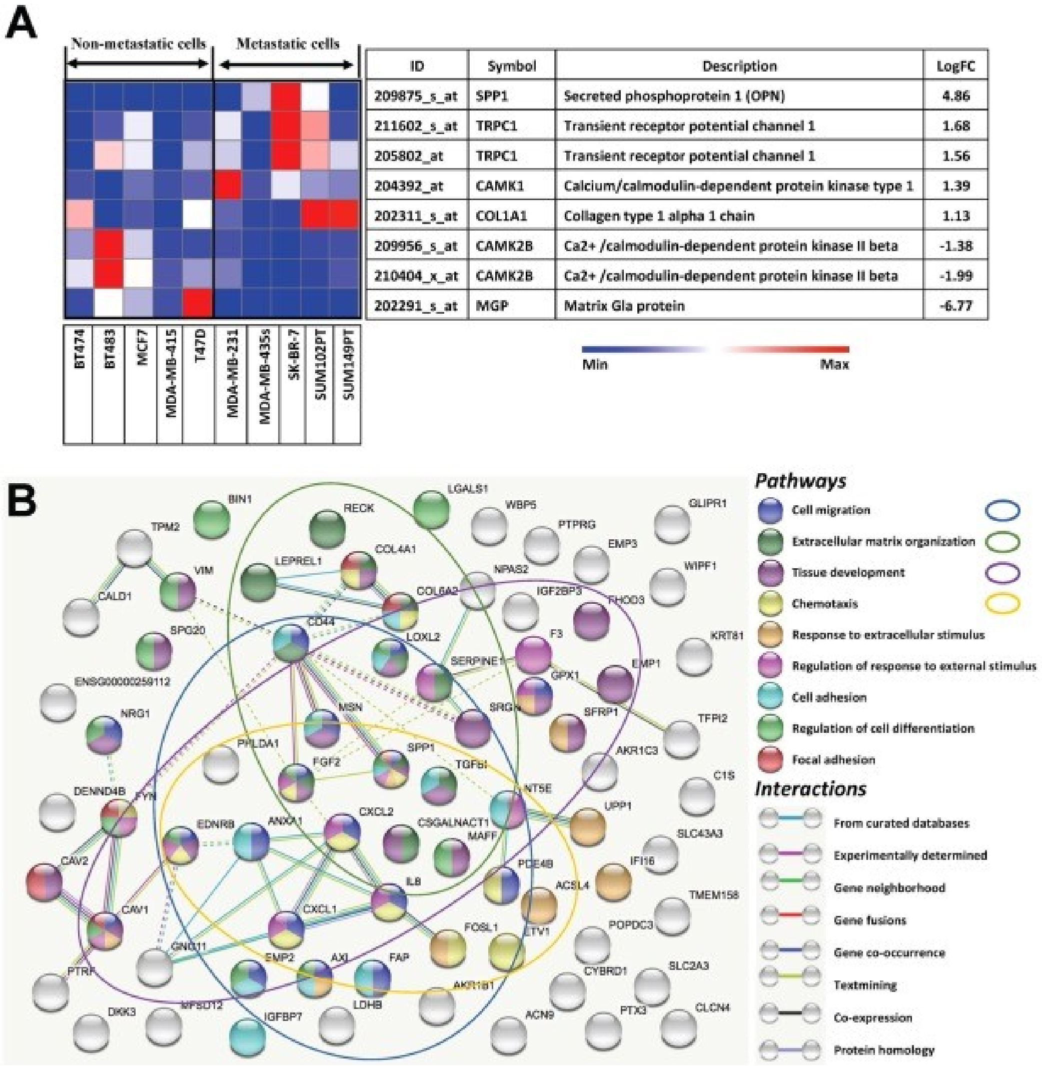
Through shRNA knockdown of OPN in human MDA-MB-231 breast cancer cells, we have shown a direct involvement of the OPN gene in the formation of microcalcifications. Moreover, OPN knockdown resulted in reduced migration in in vitro assays, which was mediated at least in part by reduced CD44. The contribution of OPN to the migratory properties of the cancer cells was validated through in vivo studies by quantifying and comparing levels of OPN and CD44 expression in parental MDA-MB-231 cells orthotopically implanted in the mouse, MDA-MB-231 cells that have escaped from the primary tumor into the blood circulation, and MDA-MB-231 cells that have successfully metastasized to the lungs.

3.2 Results

3.2.1 Osteopontin expression increases with breast cancer cell aggressiveness and osteogenic cocktail treatment

We used the GEO dataset GSE16795, which contains gene expression profiles of 39 human breast cancer cell lines and divided it into two groups of five metastatic and five non-metastatic with high relevance to our experimental work [24]. Among the genes that are differentially expressed in metastatic versus non-metastatic cell lines, the genes encoding proteins with putative roles in the context of breast microcalcifications are shown in **Fig. 3.1A** in decreasing order of their log two-fold change. **Figure 3.1A** also shows the differential expression of the listed genes across the cell lines in the dataset as a heat map. Specifically, gene expression levels of OPN were found to be significantly (p-value = 0.0047) elevated (~29 fold) for the metastatic group compared to the non-metastatic group. We analyzed the protein-protein interactions of significantly differentially expressed genes in the selected metastatic versus non-metastatic cell lines from GSE16795 using the STRING-10.5 (http://string-db.org) analysis software and database [25]. **Figure 3.1B** visualizes a subset of the identified biological processes and pathways that involve OPN (SPP1). Network nodes are colored by pathway membership, and pathways are sorted by increasing false discovery rate. Pathways that were most significantly activated, i.e. cell migration, extracellular matrix organization, tissue development, and chemotaxis, were also circled in the same color as the corresponding nodes. Additional pathways with significant activation in metastatic cells that involve OPN were response to extracellular stimulus, regulation of response to external stimulus, cell adhesion, regulation of cell differentiation, and focal adhesion. Solid and dotted lines represent intra-cluster and inter-cluster functional associations, respectively. As seen from **Fig. 3.1B**, OPN (SPP1) directly interacts with CD44 and FGF2, which in turn interact with several other proteins, including vimentin (VIM) through CD44.

As the next step, we cultured metastatic (MDA-MB-231 and SUM-149) and non-metastatic (BT-474 and T47D) human breast cancer cell lines and characterized their OPN mRNA expression levels as determined by qRT-PCR. To assess the relationship between calcification status and OPN expression level, the same cell lines were cultured in media enriched with osteogenic cocktail for induction of microcalcifications. **Figure 3.2A** shows the OPN mRNA expression results in the presence of osteogenic cocktail. The OPN mRNA expression levels of cells cultured in the absence of osteogenic cocktail are shown alongside for comparison. The OPN mRNA expression levels are significantly higher in both of the metastatic as compared to both of the non-metastatic breast cancer cell lines. It is also evident that OPN expression increases substantially with the addition of exogenous phosphates in the form of osteogenic cocktail, indicating that OPN may play a crucial role in mediating the formation of microcalcifications in breast cancer cells. The metastatic triple-negative human breast cancer cell line MDA-MB-231 was employed as the model system for further investigations in this study.



ID	Symbol	Description	LogFC
209875_s_at	SPP1	Secreted phosphoprotein 1 (OPN)	4.86
211602_s_at	TRPC1	Transient receptor potential channel 1	1.68
205802_at	TRPC1	Transient receptor potential channel 1	1.56
204392_at	CAMK1	Calcium/calmodulin-dependent protein kinase type 1	1.39
202311_s_at	COL1A1	Collagen type 1 alpha 1 chain	1.13
209956_s_at	CAMK2B	Ca2+ /calmodulin-dependent protein kinase II beta	-1.38
210404_x_at	CAMK2B	Ca2+ /calmodulin-dependent protein kinase II beta	-1.99
202291_s_at	MGP	Matrix Gla protein	-6.77

Figure 3.1: Gene expression analysis of microcalcifications in breast cancer cells. (**A**) Expression profiles of genes relevant to breast microcalcifications that are differentially expressed in metastatic and non-metastatic breast cancer cell lines. The gene expression heat map focused on genes directly linked to microcalcifications was obtained by subjecting the publicly available microarray dataset GSE16795 to the Gene-e matrix visualization and analysis platform. From the 38 cell lines included in the dataset, 5 non-metastatic and 5 metastatic cell lines with high relevance to our experimental work were identified and utilized to generate the heat map. (**B**) Protein-protein interaction network of differentially expressed genes in the selected metastatic *versus* non-metastatic cell lines. Identified biological processes and pathways that involve OPN (SPP1) are shown. Network nodes are colored by pathway membership, and pathways are sorted by increasing false discovery rate. Interactions are colored by type of interaction as

listed in the legend. Solid and dotted lines represent intra-cluster and inter-cluster functional associations, respectively.

3.2.2 Stable shRNA silencing of OPN inhibits the formation of cellular microcalcifications

To study the role of the OPN gene (SPP1) in the formation of microcalcifications in MDA-MB-231 cells, stable clones were generated using shRNA knockdown of the SPP1 gene in these cells. Four stable clones - shOPN1 through shOPN4 - were identified, characterized, and used for further studies. **Figure 3.2B** shows significantly reduced OPN mRNA expression levels in the stably OPN silenced lines versus empty-vector control cells, confirming sufficient shRNA gene knockdown. Stably silenced clones and control cells were sub-cultured in the presence of osteogenic cocktail for 7 days to induce the formation of microcalcifications. The cells were fixed and stained with alizarin red S to selectively report for the presence of microcalcifications. **Figure 3.3A** shows representative images of stably OPN silenced MDA-MB-231 clones and control cells stained with alizarin red for qualitative comparison. Three independent batches of cells were stained, and the average calcification content of the cells is shown in **Fig. 3.3B** along with standard deviations. Our observations reveal that there is a consistent inhibition of the formation of cellular microcalcifications due to OPN gene silencing in the knockdown clones. The similarity in the trend of variation in level of OPN mRNA expression and cellular calcification content across the knockdown clones further strengthens our hypothesis that the OPN gene positively regulates the formation of cellular microcalcifications.

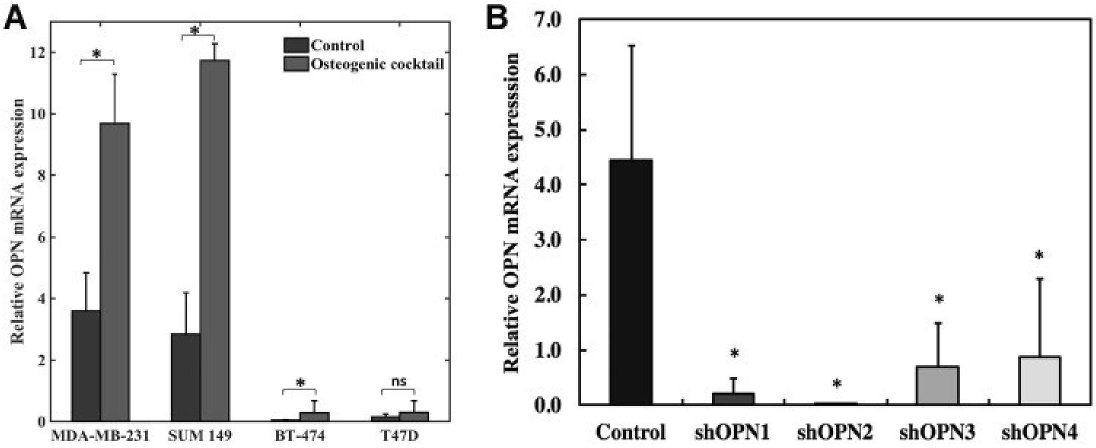


Figure 3.2: OPN mRNA expression of breast cancer cells in osteogenic cocktail. (**A**) Enhancement of OPN mRNA levels in metastatic (MDA-MB-231 and SUM 149) and non-metastatic (BT-474 and T47D) cell lines in response to addition of osteogenic cocktail. Relative expression levels of OPN mRNA in selected cell lines (analyzed using qRT-PCR) in response to osteogenic cocktail is shown. The expression levels of OPN in cells grown in control media are shown alongside for comparison. (**B**) Generation of stable clones that exhibit reduced OPN expression using shRNA silencing of the SPP1 (i.e OPN) gene in MDA-MB-231 cells. Relative expression of OPN mRNA in stably silenced clones (analyzed using qRT-PCR) in response to osteogenic cocktail is shown. Conventional Student t test threshold ($p < 0.05$) was considered statistically significant and is indicated by *.

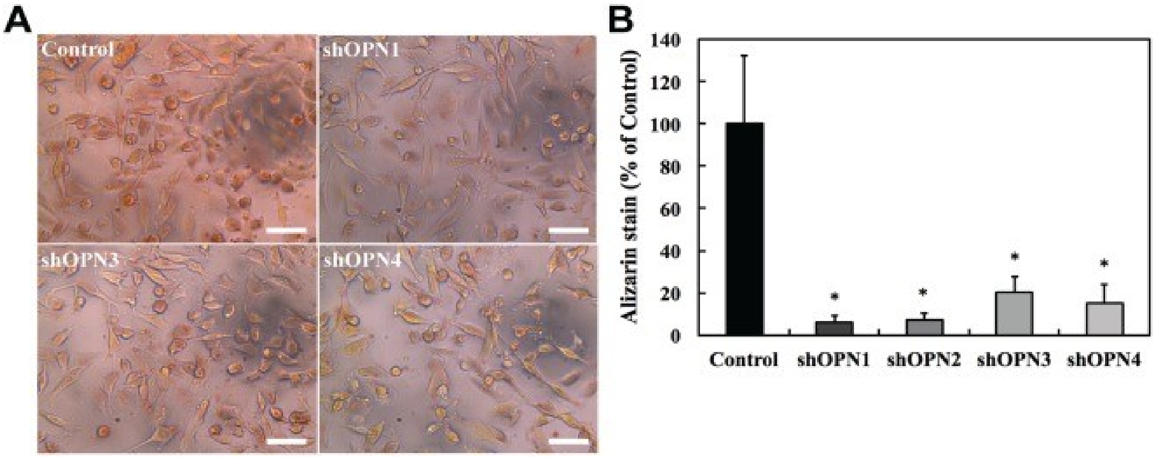


Figure 3.3: Silencing of OPN gene results in inhibition of cellular microcalcification formation in the knockdown clones. (**A**) Representative bright-field images showing alizarin red S stained cells for the four stably silenced clones and vector control. The scale bars represent 50 µm. (**B**) Bar plot showing mean and standard deviation of normalized alizarin red S stain intensity for stably silenced clones as percentage of vector control. Conventional Student t test threshold ($p < 0.05$) was considered statistically significant and is indicated by *.

3.2.3 Stable OPN silencing reduces the migration of aggressive MDA-MB-231 cells

The impact of stable OPN silencing on the migration of MDA-MB-231 cells was tested using transwell migration assays. All four stably OPN silenced MDA-MB-231 lines employed in the study, shOPN1 through shOPN4, displayed reduced cell migration compared to control cells. **Figure 3.4** shows representative migration assay images along with the quantitative comparison of migration ability for all the clones studied. The differences in migration ability were statistically significant for all clones when compared to control cells, and consistent for all the biological repeats (n = 3).

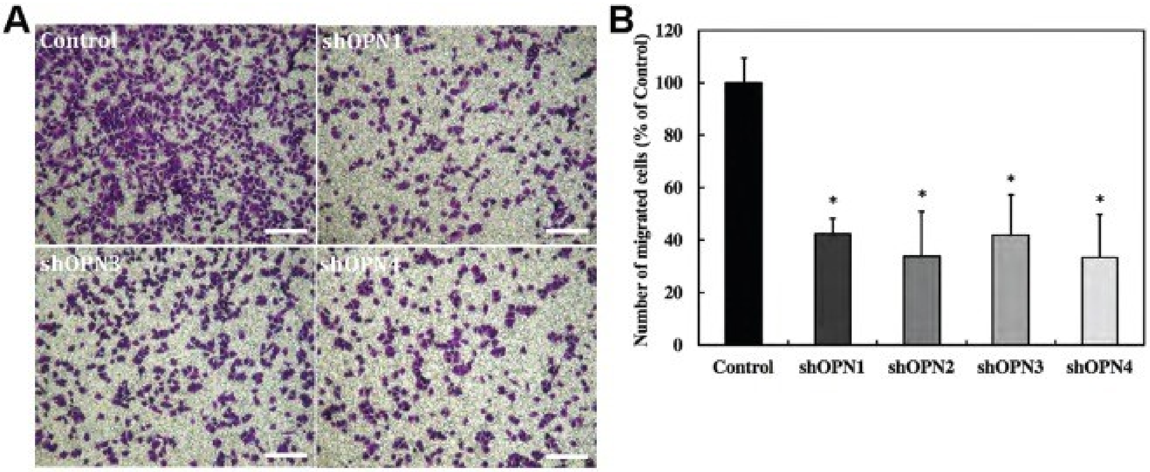


Figure 3.4: Silencing of OPN gene results in reduction of *in vitro* migration potential of the knockdown clones. (**A**) Representative bright-field images showing crystal violet stained membranes of transwell inserts for the four stably silenced clones and vector control. The scale bars represent 200 µm. (**B**) Bar plot showing mean and standard deviation of number of migrated cells for stably silenced clones as percentage of vector control. Conventional Student t test threshold ($p < 0.05$) was considered statistically significant and is indicated by *.

3.2.4 Over-expression of OPN in circulating tumor cells (CTC) and lung metastatic cells (LMC)

Motivated by the reduction of migration capabilities of the cells in the in vitro experiments, we performed orthotopic inoculations of MDA-MB-231 breast cancer cells and derived MDA-MB-231 tumor cells from the blood circulation (circulating tumor cells, CTC) and lungs (lung metastatic cells, LMC) at 8–12 weeks following inoculation as shown in **Fig. 3.5A**. These CTC and LMC were examined to verify the relevance of our findings in vivo. **Figure 3.5B** shows the mean and standard deviation of OPN mRNA expression levels comparatively for parental MDA-MB-231 cells, CTC, and LMC. The observations reveal 80-fold and 160-fold increased expression of OPN in CTC and LMC, respectively, as compared to the parental MDA-MB-231 cells. We also tested the expression levels of two of the genes identified as interacting with OPN in our protein-protein interaction analysis (**Fig. 3.1B**), which are involved in cell migration, extracellular matrix organization, and cell adhesion. The means and standard deviations of CD44 mRNA (**Fig. 3.5C**) and VIM mRNA (**Fig. 5D**) expression levels are shown comparatively for parental MDA-MB-231 cells, CTCs, and LMCs. We observed significant decreases in CD44 mRNA expression in CTCs and LMCs compared to parental MDA-MB-231 cells. VIM mRNA expression was significantly increased in CTCs and LMCs. Differences in CD44 and VIM mRNA levels between CTCs and LMCs were not significantly different.

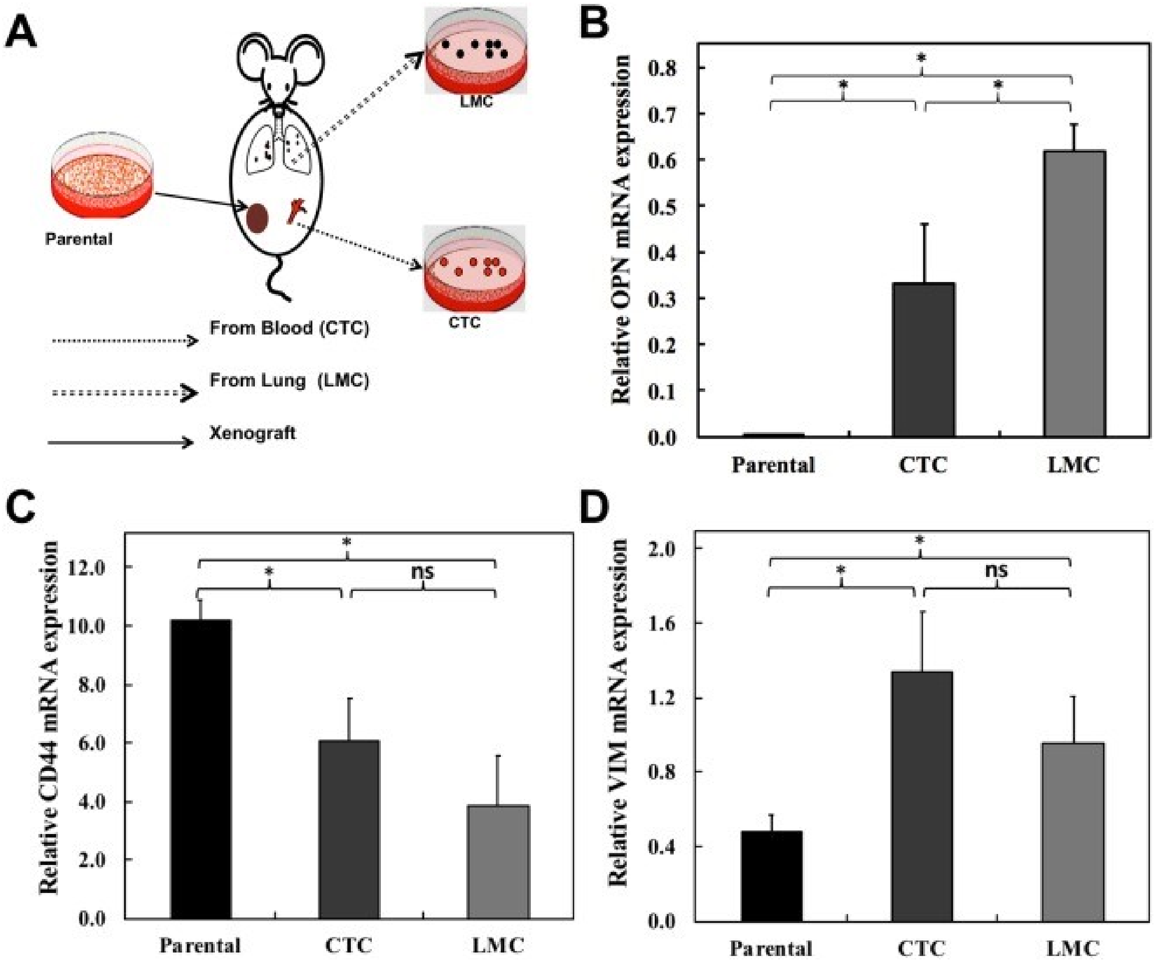


Figure 3.5: In vivo migration of MDA-MB-231 cells is dependent on expression of OPN. (**A**) Schematic of in vivo study conducted to isolate fluorescent tumor cells of varying metastatic potential from blood and metastatic lungs of mice carrying orthotopic MDA-MB-231 xenograft. Bar plots showing mean and standard deviation of (**B**) OPN mRNA, (**C**) CD44 mRNA, and (**D**) VIM mRNA expression for circulating tumor cells (CTCs) and lung metastatic cells (LMCs) relative to parental MDA-MB-231 cells. Conventional Student t test threshold ($p < 0.05$) was considered statistically significant and is indicated by *.

3.3 Discussion

Cellular mechanisms driving the formation of mammary microcalcifications in breast cancer cells remain unclear despite their extensive use in breast cancer screening and staging. Since the molecular biology of the formation of microcalcifications in breast cancer is poorly understood, we performed an initial in silico screen using an existing mRNA database for comparing highly metastatic breast cancer cell lines with non-metastatic breast cancer cell lines. In this screen, we selectively considered genes with a putative direct or indirect relationship with breast microcalcifications or bio-mineralization in general. Osteopontin, the protein encoded by the gene SPP1, which had the highest differential expression in our analysis, has previously been associated with both physiological and pathological mineralization in various organs, making it a good candidate gene for further investigation [9]. The increased expression of this gene in MDA-MB-231 cells in response to osteogenic cocktail treatment further supported the possibility of its involvement in inducing mammary mineralization. The osteogenic cocktail we used is known to induce intracellular mineralization comprising of calcium hydroxyapatite due to the increased availability of phosphate in the cell culture media with addition of β-glycerophosphate, and therefore has emerged as a popular in vitro model for studying mineralization [26].

The significant suppression of intracellular mineralization as assessed by alizarin red S staining of cells cultured in osteogenic cocktail in stably OPN silenced MDA-MB-231 cells suggests that OPN is one of the principal factors directly governing cellular mineralization processes in triple-negative breast cancer cells. Studies in the past have shown that OPN is critical for both promotion and inhibition of hydroxyapatite formation in normal bone and connective tissue cells [27,28]. More recently, it has been reported that the regulation of hydroxyapatite formation depends on the phosphorylation state of OPN and its interactions with other molecules such as osteocalcin [29]. OPN also has been reported to act as hydroxyapatite nucleator when present in certain suitable conformations [29]. Based on this evidence and the results presented in this study, we reason that the formation of microcalcifications in breast cancer cells is a result of the availability of OPN in its phosphorylated state, and its dynamic interaction with several other proteins.

Our observation that OPN regulates the formation of microcalcifications in breast cancer cells is in agreement with recent observations by Scimeca et al. that mammary microcalcifications are found in breast epithelial cells that have developed an osteoblast-like phenotype [11]. Such osteoblast-like cells developed from breast epithelial cells that were triggered by β-microglobulin to acquire mesenchymal characteristics. The osteoblast-like phenotype was sustained by bone morphogenic protein-2 (BMP-2), and these cells exhibited localized hydroxyapatite-rich cytoplasmic vesicles similar to hydroxyapatite containing intracellular vesicles found in osteoblasts [11]. In the same study, increased focal expression of OPN was observed in the proximity of hydroxyapatite, which was also a characteristic feature of lesions in biopsied tissues, showing microcalcifications along with increased mesenchymal markers such as VIM. This observation is consistent with significantly increased expression levels of the genes encoding OPN and VIM in our microarray analysis as well as in our animal model of spontaneous dissemination, in which CTCs and LMCs concurrently displayed elevated levels of OPN and VIM as compared to parental human MDA-MB-231 breast cancer cells. Together, our results support the notion that microcalcifications in aggressive breast cancer cells are driven by processes similar to that govern physiological mineralization.

Furthermore, the relationship between the microcalcification status of breast cancer cells and their metastatic capabilities remains largely unexplored. In our microarray analysis of publicly available data, we observed that OPN was significantly involved and interacted with cell migration, extracellular matrix organization, chemotaxis, and cell adhesion in metastatic breast cancer cells. Motivated by these data and the finding that microcalcifications are preferentially produced in cells displaying a mesenchymal phenotype, we assessed the effect of stable OPN knockdown on the migration capabilities of MDA-MB-231 cells. The significant inhibition of migration that we detected in all knockdown clones in in vitro transwell migration assays indicates that OPN expression directly affects the migration abilities of breast cancer cells. Our results are in good agreement with a recent study by Zhang et al., in which OPN knockdown in breast cancer cells resulted in integrin-induced inhibition of cell migration and invasion, and promoted apoptosis through induction of autophagy and inactivation of PI3K/Akt/mTOR pathway [30]. Our data also agree with recently published data showing that transient siRNA knockdown of OPN in murine mammary tumor cell lines reduced cell migration [31].

From an orthotopic mouse xenograft model of spontaneous metastasis in vivo, we observed that the OPN mRNA expression level from MDA-MB-231 cells that had spontaneously metastasized to the lungs (LMC) were significantly higher than that from circulating tumor cells (CTC) from the same model, which in turn were significantly higher than that in parental cells. These progressively increasing OPN levels in MDA-MB-231 cells of the same genetic background, which have progressively traveled farther down the metastatic cascade confirms in an in vivo system of spontaneous metastasis that dramatically increased OPN levels are most likely required for MDA-MB-231 cells to metastasize. These observations further suggest that the OPN gene is expressed differentially in the same aggressive breast cancer cell type depending on where these cancer cells are in their metastatic journey.

Concurrently, we observed progressively decreased expression levels of CD44, which were highest in parental MDA-MB-231 cells, significantly decreased in CTCs, and the most dramatically decreased in LMCs. CD44 is a cell surface glycoprotein involved in cell communication and adhesion between adjacent cells and between cells and the extracellular matrix [32]. Reduced CD44 expression levels were previously shown to enhance breast cancer metastasis [33], which is in good agreement with our observations. Our studies thus provide further proof for an interaction between OPN and CD44 that helps aggressive breast cancer cells to facilitate migration, spontaneous dissemination, and formation of metastatic nodules.

We also observed that the expression of VIM was significantly enhanced in CTCs and LMCs as compared to parental MDA-MB-231 cells. VIM is an intermediate filament protein that induces changes in the shape and motility of cells that are undergoing EMT [34,35]. Triple-negative breast cancers were previously shown to display elevated VIM expression levels compared to other types of breast tumors [36]. High expression levels of VIM in primary breast cancers were reported to support the formation of metastases in distant organs [37]. Taken together, our observations reinforce that an OPN-CD44-VIM interaction axis with implications in inducing EMT and reducing adhesion, helped triple-negative MDA-MB-231 cells to disseminate and form distant metastases.

Due to its active role in the regulation of several key pathways that have implications in disease progression, OPN is emerging as a novel therapeutic target [38,39]. The findings of the current study may enable accurate monitoring of response to such therapy through the evaluation of changes in microcalcification status. In addition, by exploring the relationship between OPN therapy and its impact on metastatic progression, changing microcalcification status can potentially be utilized as a marker to track metastatic development.

In conclusion, two major findings suggest the possibility of a fundamental relationship between mammary microcalcifications and metastatic capabilities of the cells in which they are formed. These findings are: (i) breast lesions rich in hydroxyapatite-based microcalcifications are associated with poor prognosis6, and (ii) evidence presented here that OPN is positively regulating both cellular microcalcification as well as breast cancer cell migration and metastasis, which may occur through the observed OPN-CD44-VIM axis. Future studies further exploring the causal molecular relationship between microcalcifications and metastasis in breast cancer are important for a comprehensive understanding of microcalcification etiology and their improved use as diagnostic and prognostic marker in breast cancer.

3.4 Materials and methods

3.4.1 Identification of candidate genes responsible for breast microcalcifications

The publicly available breast cancer microarray dataset GSE 16795 was analyzed where multiple breast cancer cell lines were grown to optimal cell densities for mRNA extraction and hybridization on Affymetrix microarrays [24]. A heat map was generated using the Gene-e matrix visualization and analysis platform (http://www.broadinstitute.org). This heat map represents changes in relative content of mineralization-related gene expression levels in 5 metastatic breast cancer cell lines (MDA-MB-231, MDA-MB-435s, SK-BR-7, SUM102PT, SUM149PT) and 5 non-metastatic breast cancer cell lines (BT-474, BT-483, MCF7, MDA-MB-415, T47D).

3.4.2 Protein–protein interaction network

The prominent genes overexpressed in the selected metastatic cell lines in the Gene-e analysis were employed to visualize protein-protein interactions using the STRING-10.5 (http://string-db.org) computational tool and database with a high confidence interval of 0.7 [25]. The STRING network, composed of the functional protein associations is based on genomic context, high-throughput experiments, co-expression, and scientific reports. Functional enrichments in the network were identified using the STRING tool, and a subset of the identified biological processes and pathways that involve OPN (SPP1) were selected for visualization. The nodes in the network are colored according to their membership in each of the identified pathways and the pathways are sorted in the legend by increasing false discovery rates.

3.4.3 Cell culture

The human breast cancer cell lines MDA-MB-231, SUM 149, BT-474 and T47D were obtained from the American Type Culture Collection (ATCC, MD) [40]. All cell lines were cultured in RPMI 1640 (Sigma-Aldrich) supplemented with 10% fetal bovine serum, 100 U/ml penicillin, 100 μg/ml streptomycin and 2 μg/ml fungizone antimycotic (Life Technology, Grand Island, NY, USA) and maintained in a humidified incubator in 5% CO2 at 37 °C. In a subset of the studies, the cell culture media was supplemented with an osteogenic cocktail containing 10 mM β-glycerophosphate (Sigma-Aldrich) and 50 mg/ml-1 ascorbic acid (Sigma-Aldrich) for promoting the formation of microcalcifications.

3.4.4 Generation of stably OPN silenced breast cancer cell lines

MDA-MB-231 cells were transfected with lentiviral particles expressing shRNA against human OPN (sc-36129-V, Santa Cruz Biotechnology, Dallas, Texas) to specifically knockdown the expression of the human OPN gene. These OPN shRNA lentiviral particles were purchased as a pool of concentrated, transduction-ready viral particles containing 3 target-specific constructs that encode 19–25 nt (plus hairpin) shRNA designed to knock down OPN gene expression. Stably transduced clones were developed, along with a vector control cell line expressing a control shRNA lentiviral particle (sc-108080, Santa Cruz Biotechnology). qRT-PCR to measure OPN mRNA and immunoblotting of OPN confirmed successful transduction. For stably expressing lines, transfected cells were passaged and maintained in media containing Puromycin dihydrochloride (sc-108071, Santa Cruz Biotechnology). Cells were kept under selection for 7–10 days. Then, the cells from the selection step were plated at a density of 10 cells per ml in a 96-well tissue culture plate by adding 100 µl per well (i.e., 1 cell per well). Selected single-colony wells in the 96-well tissue culture plate were expanded to high confluence and transferred to a 24-well tissue culture plate. Once the colonies in 24-well tissue culture plates were expanded to high confluence, they were passaged to a 6-well tissue culture plate. Clonal cell lines were assessed by qRT-PCR to select lines with significantly decreased levels of OPN gene expression.

3.4.5 qRT-PCR

Three cell lines from each group (MDA-MB-231, CTC, LM) were analyzed for gene expression with two technical repeats and two biological repeats each. For RNA purification, cells were grown for 48 hours in exponential growth phase and mRNA was isolated and purified using the RNeasy total RNA isolation kit (Qiagen, Germantown, MD) according to the manufacturer’s protocol. mRNA was reverse transcribed into cDNA using qScript™ cDNA SuperMix (Quanta Bioscience, Gaithersburg, MD). Quantitative real-time PCR (q-RT-PCR) was performed using IQ SYBR Green Supermix and gene-specific primers in the iCycler RT-PCR detection system (Bio-Rad, Hercules, CA) with 2 µl of diluted cDNA samples (1:10) used as a template using the following primers. The sequences for forward and reverse primers for quantifying SPP1 mRNA, which expresses the OPN protein, were 5′-CGAGGTGATAGTGTGGTTTATGG-3′ and 5′-GCACCATTCAACTCCTCGCTTTC-3′, respectively. The sequences for forward and reverse primers for quantifying CD44 mRNA were 5′-CGGACACCATGGACAAGTTT-3′ and 5′-GAAAGCCTTGCAGAGGTCAG-3′, respectively. The sequences for forward and reverse primers for quantifying VIM mRNA were 5′-GCAAAGATTCCACTTTGCGT-3′ and 5′-GAAATTGCAGGAGGAGATGC-3′, respectively. The housekeeping gene Hypoxanthinephophoribosyltransferase-1 (HPRT1) was used as internal reference gene for quantification [41] as previously described [42,43]. The expression of RNA relative to HPRT1 was calculated [42–44] based on the Ct as $R = 2^{-(\Delta Ct)}$, where $\Delta Ct = Ct_{target} - Ct_{HPRT1}$ mRNA gene expression level was reported as mean ± standard deviation.

3.4.6 Alizarin Red S staining and quantification of mineralization

The cell monolayers were fixed using 4% formaldehyde after washing gently with PBS. Alizarin red S staining solution at pH 4.1–4.3 was added to the fixed cell monolayers and incubated in the dark for 45 min. Cell monolayers were carefully washed with distilled water and PBS after aspirating the remaining alizarin red S solution. The stained cells were imaged using a camera (Lumenera Infinity 1) mounted on a microscope (Leica DMIL, 0.4 NA and 20 × magnification objective). Several images of the stained cells were captured from three independent cultures for each stable clone, (each frame capturing more than 50 cells) and the pixels corresponding to alizarin red were quantified and normalized to the total number of cells per frame to remove the effects of any differences in cell densities across the clones. The pixels corresponding to the alizarin red stained areas were identified by their RGB decomposition obtained using MATLAB (Mathworks, Natick, MA) and the cells in each frame were counted using the manual mode of Image-J software [45]. The colors determined by the criterion $R > 180$, $R > G + 80$ and $B < 100$ were found to accurately represent the color of the alizarin red S stain.

3.4.7 In Vitro migration assays

Transwell inserts (Costar) with porous polycarbonate membranes with a pore size of 8 µm were used to measure the effect of stable OPN silencing on migration in MDA-MB-231 cells. shRNA knockdown and control clones (1×10^5 cells) suspended in 100 µL of serum-free RPMI were added to the upper chamber of the insert and allowed to migrate across the membranes, which occurred under the influence of RPMI medium with 5% fetal bovine serum as chemoattractant in the lower chamber. After 20 hours, the lower sides of the membranes were fixed in 4% formalin and stained with 0.2% crystal violet solution. After staining, four fields of view for each insert were obtained with an inverted microscope at 10 × magnification. Quantitative measurements of the number of cells migrating across the membrane were obtained by applying an intensity threshold after converting the RBG images to 8-bit grayscale images using Image-J software [45].

3.4.8 Experimental animal models: Generation of CTC and LMC

All animal experiments were approved by the Johns Hopkins University Animal Care and Use Program in compliance with the Animal Welfare Act regulations and

Public Health Service (PHS) policy. Johns Hopkins University has an approved PHS Assurance and maintains accreditation of our program by the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC) International. MDA-MB-231 cells were stably transfected with a construct containing DNA of tdTomato fluorescent protein as previously described [43]. Stably tdTomato-expressing MDA-MB-231 breast cancer cells (2×10^6) were orthotopically implanted into the fourth right mammary fat pad of 6 weeks old female athymic nu/nu female mouse (NCI) as described previously [42,43,46]. When primary tumor volume reached approximately 500 mm^3 after about 8–12 weeks following inoculation, the mouse was sacrificed, its blood was obtained by cardiac puncture, and its lungs were collected to isolate and culture CTC and LMC, respectively, as briefly described in the following. For CTC isolation, cardiac puncture yielded about 500 µl of blood from each mouse. Red blood cells were lysed (ACK lysing buffer, Life technology, Grand Island, NY, USA) and CTCs were pelleted by centrifugation, suspended in RPMI 1640 culture medium, and CTC presence was verified by fluorescence microscopy of tdTomato-expression in MDA-MB-231 cells. For LMC isolation, both lungs were carefully removed and cut into 4 mm sized tissue pieces with a sterile scalpel and scissors. These lung tissue pieces were placed onto sterile Pyrex petri dishes, washed 3 times in balanced salt solution without calcium and magnesium, and finely chopped in 0.25% trypsin/EDTA solution. Lung tissue was digested at 37 °C for 1 hour to maximize trypsin penetration. Warm, complete media was added to the tissue pieces, which was gently dispersed by pipetting. The resulting tissue suspension was then passed twice through a 20 G syringe needle to completely disperse any remaining tissue. Two weeks after cell culture of CTC and LMC, the tdTomato-expressing CTC or LMC cells were sorted by FACS to clear out all non-fluorescent non-cancer cells of lung origin.

3.4.9 Quantification and statistical analysis

Statistical significance of the differences between quantitative measurements were analyzed by unpaired two-tailed Student's t-test. P-values < 0.05 were considered to be statistically significant.

Bibliography

[1] R. L. Siegel, K. D. Miller and A. Jemal, Cancer statistics, 2018, *CA: A Cancer Journal for Clinicians* **68**(1), 2018, 7-30.

[2] M. Gülsün, F. B. Demirkazık and M. Arıyürek, Evaluation of breast microcalcifications according to breast imaging reporting and data system criteria and le gal's classification, *European Journal of Radiology* **47**(3), 2003, 227-231.

[3] J. Evans, A. R. M. Wilson, H. C. Burrell, I. O. Ellis and S. E. Pinder, Mammographic features of ductal carcinoma in situ (dcis) present on previous mammography, *Clinical Radiology* **54**(10), 1999, 644-646.

[4] J. M. Johnson, R. R. Dalton, S. M. Wester, J. Landercasper and P. J. Lambert, Histological correlation of microcalcifications in breast biopsy specimens, *Archives of Surgery* **134**(7), 1999, 712-716.

[5] M. Morgan, M. Cooke and G. Mccarthy, Microcalcifications associated with breast cancer: An epiphenomenon or biologically significant feature of selected tumors?, *J Mammary Gland Biol Neoplasia* **10**(2), 2005, 181-187.

[6] R. Sathyavathi, A. Saha, J. S. Soares, N. Spegazzini, S. Mcgee, R. Rao Dasari, et al., Raman spectroscopic sensing of carbonate intercalation in breast microcalcifications at stereotactic biopsy, *Sci Rep* **5**(2015, 9907.

[7] T. Kirsch, Determinants of pathological mineralization, *Current Opinion in Rheumatology* **18**(2), 2006, 174-180.

[8] R. C. Shroff and C. M. Shanahan, Vascular calcification in patients with kidney disease: The vascular biology of calcification, *Seminars in Dialysis* **20**(2), 2007, 103-109.

[9] A. Bellahcene and V. Castronovo, Increased expression of osteonectin and osteopontin, two bone matrix proteins, in human breast cancer, *The American journal of pathology* **146**(1), 1995, 95-100.

[10] A. Bellahcène, M. P. Merville and V. Castronovo, Expression of bone sialoprotein, a bone matrix protein, in human breast cancer, *Cancer Research* **54**(11), 1994, 2823-2826.

[11] M. Scimeca, E. Giannini, C. Antonacci, C. Pistolese, L. Spagnoli and E. Bonanno, Microcalcifications in breast cancer: An active phenomenon mediated by epithelial cells with mesenchymal characteristics, *BMC Cancer* **14**(1), 2014, 286.

[12] D. R. Senger, C. A. Perruzzi, A. Papadopoulos and D. G. Tenen, Purification of a human milk protein closely similar to tumor-secreted phosphoproteins and osteopontin, *Biochimica et biophysica acta* **996**(1-2), 1989, 43-48.

[13] K. X. Wang and D. T. Denhardt, Osteopontin: Role in immune regulation and stress responses, *Cytokine & growth factor reviews* **19**(5-6), 2008, 333-345.

[14] D. R. Senger, D. F. Wirth and R. O. Hynes, Transformed mammalian cells secrete specific proteins and phosphoproteins, *Cell* **16**(4), 1979, 885-893.

[15] H. Gardner, B. Berse and D. R. Senger, Specific reduction in osteopontin synthesis by antisense rna inhibits the tumorigenicity of transformed rat1 fibroblasts, *Oncogene* **9**(8), 1994, 2321-2326.

[16] E. I. Behrend, A. M. Craig, S. M. Wilson, D. T. Denhardt and A. F. Chambers, Reduced malignancy of ras-transformed nih 3t3 cells expressing antisense osteopontin rna, *Cancer Research* **54**(3), 1994, 832.

[17] H. Singhal, D. S. Bautista, K. S. Tonkin, F. P. O'malley, A. B. Tuck, A. F. Chambers, et al., Elevated plasma osteopontin in metastatic breast cancer associated with increased tumor burden and decreased survival, *Clinical Cancer Research* **3**(4), 1997, 605-611.

[18] L. R. Rodrigues, J. A. Teixeira, F. L. Schmitt, M. Paulsson and H. Lindmark-Mansson, The role of osteopontin in tumor progression and metastasis in breast cancer, *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**(6), 2007, 1087-1097.

[19] N. Y. Li, C. E. Weber, Z. Mi, P. Y. Wai, B. D. Cuevas and P. C. Kuo, Osteopontin up-regulates critical epithelial-mesenchymal transition transcription factors to induce an aggressive breast cancer phenotype, *Journal of the American College of Surgeons* **217**(1), 2013, 17-26; discussion 26.

[20] Y. Zhu, X. M. Gao, J. Yang, D. Xu, Y. Zhang, M. Lu, et al., C-c chemokine receptor type 1 mediates osteopontin-promoted metastasis in hepatocellular carcinoma, *Cancer science* 2017,

[21] G. F. Weber, S. Ashkar, M. J. Glimcher and H. Cantor, Receptor-ligand interaction between cd44 and osteopontin (eta-1), *Science* **271**(5248), 1996, 509-512.

[22] M. Ahmed, J. L. Sottnik, G. M. Dancik, D. Sahu, D. E. Hansel, D. Theodorescu, et al., An osteopontin/cd44 axis in rhogdi2-mediated metastasis suppression, *Cancer cell* **30**(3), 2016, 432-443.

[23] Y. U. Katagiri, J. Sleeman, H. Fujii, P. Herrlich, H. Hotta, K. Tanaka, et al., Cd44 variants but not cd44s cooperate with β1-containing integrins to permit cells to bind to osteopontin independently of arginine-glycine-aspartic acid, thereby stimulating cell motility and chemotaxis, *Cancer Research* **59**(1), 1999, 219-226.

[24] A. Hollestelle, J. H. Nagel, M. Smid, S. Lam, F. Elstrodt, M. Wasielewski, et al., Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines, *Breast Cancer Res Treat* **121**(1), 2009, 53-64.

[25] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, et al., String v10: Protein–protein interaction networks, integrated over the tree of life, *Nucleic Acids Research* **43**(D1), 2015, D447-D452.

[26] R. F. Cox, A. Hernandez-Santana, S. Ramdass, G. Mcmahon, J. H. Harmey and M. P. Morgan, Microcalcifications in breast cancer: Novel insights into

the molecular mechanism and functional consequence of mammary mineralisation, *Br J Cancer* **106**(3), 2012, 525-537.

[27] A. L. Boskey, M. Maresca, W. Ullrich, S. B. Doty, W. T. Butler and C. W. Prince, Osteopontin-hydroxyapatite interactions in vitro: Inhibition of hydroxyapatite formation and growth in a gelatin-gel, *Bone and Mineral* **22**(2), 1993, 147-159.

[28] M. D. Mckee and A. Nanci, Osteopontin: An interfacial extracellular matrix protein in mineralized tissues, *Connective Tissue Research* **35**(1-4), 1996, 197-205.

[29] A. Gericke, C. Qin, L. Spevak, Y. Fujimoto, W. T. Butler, E. S. Sørensen, et al., Importance of phosphorylation for osteopontin regulation of biomineralization, *Calcified tissue international* **77**(1), 2005, 45-54.

[30] H. Zhang, M. Guo, J. H. Chen, Z. Wang, X. F. Du, P. X. Liu, et al., Osteopontin knockdown inhibits α_v,β₃ integrin-induced cell migration and invasion and promotes apoptosis of breast cancer cells by inducing autophagy and inactivating the pi3k/akt/mtor pathway, *Cellular Physiology and Biochemistry* **33**(4), 2014, 991-1002.

[31] S. Saleh, D. E. Thompson, J. Mcconkey, P. Murray and R. A. Moorehead, Osteopontin regulates proliferation, apoptosis, and migration of murine claudin-low mammary tumor cells, *BMC Cancer* **16**(1), 2016, 359.

[32] M. Zöller, Cd44: Can a cancer-initiating cell profit from an abundantly expressed molecule?, *Nature Reviews Cancer* **11**(2011, 254.

[33] J. I. Lopez, T. D. Camenisch, M. V. Stevens, B. J. Sands, J. Mcdonald and J. A. Schroeder, Cd44 attenuates metastatic invasion during breast cancer progression, *Cancer Research* **65**(15), 2005, 6755.

[34] M. G. Mendez, S.-I. Kojima and R. D. Goldman, Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition, *The FASEB Journal* **24**(6), 2010, 1838-1851.

[35] Y. Takeyama, M. Sato, M. Horio, T. Hase, K. Yoshida, T. Yokoyama, et al., Knockdown of zeb1, a master epithelial-to-mesenchymal transition (emt) gene, suppresses anchorage-independent cell growth of lung cancer cells, *Cancer Letters* **296**(2), 2010, 216-224.

[36] P. Karihtala, P. Auvinen, S. Kauppila, K.-M. Haapasaari, A. Jukkola-Vuorinen and Y. Soini, Vimentin, zeb1 and sip1 are up-regulated in triple-negative and basal-like breast cancers: Association with an aggressive tumour phenotype, *Breast Cancer Research and Treatment* **138**(1), 2013, 81-90.

[37] C. Alix-Panabières and K. Pantel, Challenges in circulating tumour cell research, *Nature Reviews Cancer* **14**(2014, 623.

[38] M. Bandopadhyay, A. Bulbule, R. Butti, G. Chakraborty, P. Ghorpade, P. Ghosh, et al., Osteopontin as a therapeutic target for cancer, *Expert Opin Ther Targets* **18**(8), 2014, 883-895.

[39] B. Tuck, A. F. Chambers and A. L. Allan, Osteopontin overexpression in breast cancer: Knowledge gained and possible implications for clinical management, *J Cell Biochem* **102**(4), 2007, 859-868.

[40] R. M. Neve, K. Chin, J. Fridlyand, J. Yeh, F. L. Baehner, T. Fevr, et al., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes, *Cancer cell* **10**(6), 2006, 515-527.

[41] J. B. De Kok, R. W. Roelofs, B. A. Giesendorf, J. L. Pennings, E. T. Waas, T. Feuth, et al., Normalization of gene expression measurements in tumor tissues: Comparison of 13 endogenous control genes, *Laboratory investigation; a journal of technical methods and pathology* **85**(1), 2005, 154-159.

[42] A. Rizwan, C. Bulte, A. Kalaichelvan, M. Cheng, B. Krishnamachary, Z. M. Bhujwalla, et al., Metastatic breast cancer cells in lymph nodes increase nodal collagen density, *Sci Rep* **5**(2015, 10002.

[43] A. Rizwan, M. Cheng, Z. M. Bhujwalla, B. Krishnamachary, L. Jiang and K. Glunde, Breast cancer cell adhesome and degradome interact to drive metastasis, *Npj Breast Cancer* **1**(2015, 15017.

[44] T. D. Schmittgen and K. J. Livak, Analyzing real-time pcr data by the comparative c(t) method, *Nature protocols* **3**(6), 2008, 1101-1108.

[45] A. Schneider, W. S. Rasband and K. W. Eliceiri, Nih image to imagej: 25 years of image analysis, *Nat Meth* **9**(7), 2012, 671-675.

[46] A. Rizwan, I. Serganova, R. Khanin, H. Karabeber, X. Ni, S. B. Thakur, et al., Relationships between ldh-a, lactate and metastases in 4t1 breast tumors, *Clin Cancer Res* **19**(18), 2013, 5158-5169.

Chapter 4

Label-free Raman spectroscopy detects stromal adaptations in pre-metastatic lungs

Recent advances in animal modeling, imaging technology and functional genomics have permitted precise molecular observations of the metastatic process. Yet, a comprehensive understanding of the pre-metastatic niche remains elusive owing to the limited set of tools that can map subtle differences in molecular mediators in organ-specific microenvironments. The challenge of measuring multiple cell types in secondary sites, prior to the arrival of metastasizing tumor cells and the consequent morphological manifestations, demands non-perturbative analytical methods that can offer quantitative biochemical insights. Here we report an approach to detect pre-metastatic changes in the lung microenvironment, in response to primary breast tumors, using a combination of metastatic mouse models, Raman spectroscopy and multivariate analysis of consistent patterns in molecular expression. We used tdTomato fluorescent protein expressing MDA-MB-231 and MCF-7 cells, of high and low metastatic potential respectively, to grow orthotopic xenografts in athymic nude mice and allow spontaneous dissemination from the primary mammary fat pad tumor. Label-free Raman spectroscopic

---

This chapter is a reprint of the peer-reviewed publication: Paidi SK*, Rizwan A*, Zheng C*, Cheng M, Glunde K, Barman I. “Label-free Raman spectroscopy detects stromal adaptations in pre-metastatic lungs primed by breast cancer”, Cancer Research, 77(2), 247-56, 2017. (* denotes equal contribution) mapping was employed to record the molecular content of pre-metastatic lungs. Our measurements show reliable distinctions in vibrational features, characteristic of the collageneous stroma and its cross-linkers as well as proteoglycans, that uniquely identify the metastatic potential of the primary tumor by recapitulating the compositional changes in the lungs. Consistent with histological assessment of the tissue specimen and gene expression analysis, our study suggests that remodeling of the extracellular matrix components as indicated by their vibrational spectroscopic signatures may present promising markers for objective recognition of the pre-metastatic niche, independent of conventional clinical information.

4.1 Introduction

While local breast cancers are largely responsive to current therapeutic strategies, treatments to permanently eradicate metastasis are yet to be developed. Consequently, nearly all breast cancer-related deaths today result from metastatic disease that involves distant organs [1]. The distribution of metastases is a non-random process with each tumor type manifesting a characteristic pattern of metastatic involvement in distant vital organs [2, 3]. Stephen Paget's 'Seed and Soil' hypothesis originally shifted the attention from a sole focus on the behavior of primary tumor cells to the important role of the stroma at the secondary site [4, 5]. Seeking to understand the basis of metastasis organotropism, his seminal hypothesis postulated that a receptive microenvironment at the secondary organ (soil) is crucial to the engraftment of circulating tumor cells (seed). This also provided a conceptual framework for later observations in experimental metastasis assays that cancer cells derived from a distant site display enhanced metastatic ability to that specific organ [6]. Yet, it is only with recent advances in animal metastasis assays, genomic profiling and real-time imaging techniques that the molecular components that drive organ-specific metastasis have been specifically probed. Translation of the preclinical findings on the metastatic microenvironment into a clinical test, however, has not yet been realized.

Building on the seed and soil hypothesis, emerging evidence suggests the formation of a pre-metastatic niche [7, 8], i.e. collective changes at the target metastasis sites prior to the arrival of the first tumor cells. This niche development in the preferred metastatic sites appears to be driven by soluble growth factors secreted by the primary tumor and recruitment of tumor associated cells [9]. The priming of the secondary organs was initially attributed to the localization of haematopoietic bone marrow progenitor cells expressing vascular epithelial growth factor receptor 1 (VEGFR-1) due to VEGF being secreted by the primary tumor [7]. Exosomes secreted from primary tumors have also been reported to play a significant role in mobilizing these progenitor cells to the pre-metastatic sites [10]. The recruitment of tumor-associated cells provides an increased availability of chemokines, growth factors, matrix degrading factors and adhesion molecules that initiate the metastatic cascade [8, 9]. This process is reported to be accompanied by remodeling of the extracellular matrix (ECM) in the pre-metastatic niche, notably through the upregulated expression of matrix metalloproteinases (MMPs) [11], transformation of local fibroblasts, and focal expression of fibronectin. For instance, a recent series of investigations have revealed that lysyl oxidase (LOX), an enzyme secreted by hypoxic tumor cells, modulates the ECM in pre-metastatic sites by crosslinking collagen fibrils, thereby making it more receptive to further myeloid cell infiltration [12, 13].

While promising, these findings also highlight the need for further research to reveal a holistic picture of the pre-metastatic stage that trigger (or inhibit) engraftment and proliferation. This, in turn, demands molecular-specific and quantitative analytical tools that can provide direct readouts from multiple biomolecules without necessitating individual labeling. Such a tool would inform if and how the compositional contributors of the stromal microenvironment in metastatic sites are changing in response to a spontaneously disseminating primary tumor - but prior to the arrival of tumor cells. Vibrational spectroscopy offers a promising tool to meet these demands, owing to the wealth of intrinsic molecular information (that obviates the need for imaging probes), extensive multiplexing capability and facile readout [14-17].

Spontaneous Raman spectroscopy, in particular, has emerged as an attractive technique for the diagnosis of cancers with high specificity and free of inter-observer variability [18]. Based on inelastic scattering of light arising from the interactions with the tissue being analyzed, Raman spectroscopy affords sub-cellular signal localization and can easily be extended to in vivo approaches [19, 20]. Recently, its ability to discern pathologies in advance of their clinical manifestations has also been shown [21]. Malins and co-workers elegantly demonstrated the early detection sensitivity of vibrational spectroscopy in a study, where spectral changes in the DNA of primary tumor were noted 57 days prior to the appearance of histological changes [22]. We hypothesized that the utility of

Raman spectroscopic information could also be extended to identifying the pre-metastatic niche, due to the unique structural and chemical changes associated with the evolving soil.

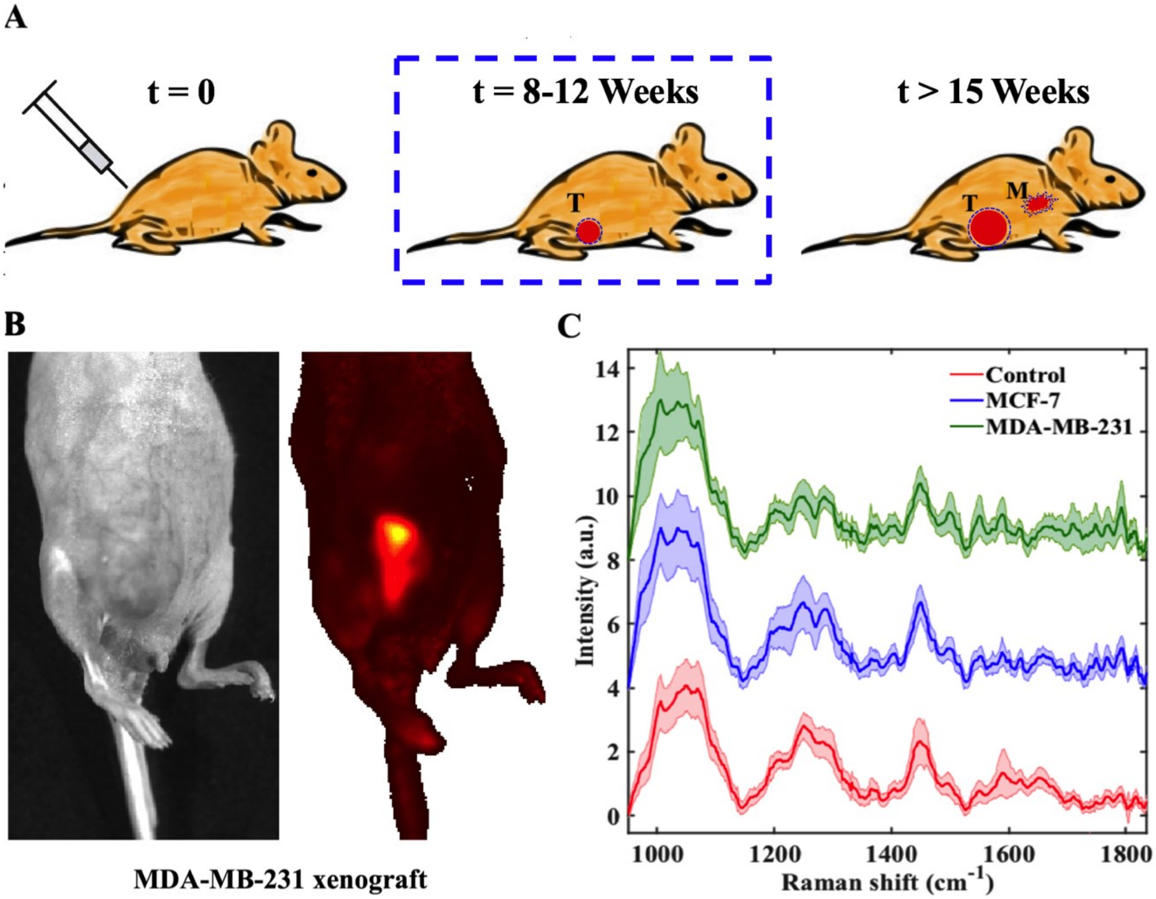


Figure 4.1: Raman spectroscopic profiling of pre-metastatic lungs. (**A**) Mouse models, orthotopically xenografted with human breast cancer cells of different metastatic potential (MCF-7 and MDA-MB-231), were used to study stromal adaptations in the lung, prior to seeding of tumor cells. (**B**) Representative *in vivo* brightfield (left) and fluorescence (right) images of mouse growing a tdTomato-expressing breast tumor xenograft. (**C**) Mean Raman spectra (with the shadow representing ±1 standard deviation) acquired from lungs of normal mice, and pre-metastatic lungs of MCF-7 and MDA-MB-231 xenografted mice are shown.

Important clues also come from a recent report by Kwak et al., demonstrating the utility of infrared (IR) spectroscopic imaging in predicting cancer recurrence by exploiting molecular features of the tumor microenvironment [23], and our recent observation that lymph nodes in mice with metastatic tumor xenografts displayed an increased collagen I density [24]. Consistent with these recent literature reports, we suspected that the collagen architectural modifications, in part, preceded the seeding of metastatic cancer cells. Because Raman spectra report vibrational features characteristic of collagen and its cross-linking moieties as well as glycoproteins, our goal in this study was to identify Raman spectral patterns that are able to detect characteristic molecular changes in the pre-metastatic niche.

Here we have investigated lungs from mouse models that recapitulate spontaneously disseminating breast cancer cells of low and high metastatic potential, and exploited the molecular basis of Raman spectroscopy to probe the pre-metastatic niche (**Figure 4.1**). Raman spectroscopic mapping measurements revealed subtle, but consistent, changes in the vibrational features of ECM components of the lungs, in particular in their collagen fiber matrix and proteoglycan content. The definition of the pre-metastatic adaptations in spectral terms facilitated the development of a decision algorithm, which accurately differentiates lungs in mice with metastatic MDA-MB-231 tumor xenografts from that in mice with MCF-7 xenografts and normal controls. A continuous model of ECM modifications, based on the metastatic potential of the primary tumor, is proposed to explain the differential signatures – in the confirmed absence of any tumor cells in the lungs. This model is in agreement with observations from Masson's trichrome staining and gene expression analysis performed on microarray data of pre-metastatic lung samples from mice harboring breast tumor xenografts. Taken together, this study demonstrates the potential of Raman spectroscopy as a rapid, objective and label-free tool in the recognition of pre-metastatic changes. We envision that our findings here will also accelerate the use of Raman spectroscopy in identifying distinct biochemical signatures in organ-specific niches, thereby enabling a better understanding of organotropism.

4.2 Materials and Methods

4.2.1 Tissue preparation and histopathology

Six-week-old female athymic nu/nu mice (NCI, MD) were orthotopically inoculated with $2x10^6$ cells of the human breast cancer cell lines MDA-MB-231 (n=3), or MCF-7 (n=3) in their fourth right mammary fat pad, as detailed in our previous article [25]. For comparison, control mice (n=3) without tumor cell implantation were employed in the study. Cell lines were obtained from the American Type Culture Collection (ATCC, MD) and stably transfected with a construct containing cDNA of tdTomato as described in our previous report [24]. Cell lines tested negative for mycoplasma and were authenticated using short tandem repeat (STR) profiling prior to inoculation in mice. Cell lines were maintained in RPMI 1640 (Sigma Aldrich) supplemented with 10% fetal bovine serum (Sigma Aldrich) and 1% penicillin-streptomycin (Sigma Aldrich) in a humidified incubator at $37^oC/5\%\ CO_2$. Prior to implantation of MCF-7 cells, mice were supplemented with 17β-Estradiol (Innovative Research of America, SE#121, 0.72 mg/pellet, 60 day release) in their neck region [26]. Primary tumor size was monitored, and mice were sacrificed within 8-12 weeks of cell implantation when primary tumors grew to *ca.* 500 mm^3 in volume. Control mice were also sacrificed in this timeframe. Freshly excised lungs of mice were cleaned in phosphate buffered saline (PBS) and fixed in formalin for 24 hours. Formalin fixed lung tissue samples were rinsed thoroughly in excess PBS to remove any residual formalin before acquiring Raman spectra. Following spectral acquisition, tissues were stored in 70% ethanol and sent to JHU Histology Services for paraffin embedding and serial sectioning, after which one of the sections was used for haematoxylin and eosin (H&E) staining. The unstained slides were utilized in our laboratory to perform Masson's trichrome staining for collagen as detailed in our previous study [24]. The Institutional Animal Care and Use Committee at the Johns Hopkins University School of Medicine approved the protocol of this study.

4.2.2 Acquisition of Raman spectra

Formalin fixed lung specimens were rinsed in PBS, flattened and placed on a clean aluminum block. There was no interference of the tissue Raman spectrum from the aluminum substrate, which also ensured a consistent probe-tissue imaging distance. A custom-built portable, fiber-probe based Raman spectroscopy system was used for spectral acquisition [27]. Briefly, an 830 nm diode laser (500 mW maximum power, Process Instruments, Salt Lake City, UT) was used to excite the sample. A lensed fiber-optic Raman bundled contact probe (Emvision LLC, FL) having a diameter of 2 mm (and an estimated tissue sampling volume of 1 mm^3) was used to deliver the excitation beam though its central fiber and collect the back-scattered light through an annular ring of optical fibers. The scattered light was directed to a spectrograph (Holospec f/1.8i, Kaiser Optical Systems, MI). The spectra were then recorded using a thermoelectrically cooled CCD camera (PIXIS 400BR, 20×20 µm pixels, 1340×400 array, Princeton Instruments, NJ). The laser power at the lung tissue samples was maintained at around 15 mW in this study and the tissue was kept moist throughout the period of laser exposure by intermittent addition of PBS. A total collection time of 10 seconds (10 accumulations of 1s each to prevent CCD saturation) was used for acquisition of each spectrum. Spectroscopic mapping was pursued to overcome the limitations of conventional fiber probe-based point spectroscopy that only examines a small area of tissue and suffers from undersampling. Wide area mapping, over the entire lung specimen, was performed by scanning the optical probe using a pair of motorized translation stages (travel range: 13 mm, T-LS13M, Zaber Technologies Inc, Vancouver, Canada) in each orthogonal direction. Zaber console (open-source software) was employed to control the raster scan through the PC serial ports. The mapping protocol also ensured the collection of sufficient spectra (approx. 300 spectra per mouse) for the development of robust classification models.

4.2.3 Data analysis

The Raman instrument was wavenumber-calibrated using 4-acetamidophenol (Tylenol©) spectra. Raman spectra recorded from mouse lungs were restricted to the fingerprint wavenumber region (500-1850 cm^{-1}) for analysis and normalized to lie between 0 and 1 in order to remove the effects of potential differences in laser power at the sample. Principal component analysis (PCA) was used to reduce the dimensionality of the spectral dataset to a few dimensions characteristic of the maximum variance in the dataset [28]. This transformation converts the set of spectral recordings into a set of values of linearly uncorrelated variables that form an orthogonal basis set. The spectral dataset of each mouse model was subjected to PCA using the statistical toolbox of MATLAB 2015b (Mathworks, Natick, MA) to obtain principal component (PC) scores and loadings that highlight the spectral features characteristic of the class. The use of these key patterns (PCs) enhances sensitivity of the analysis by not focusing on small differences in Raman signatures that may arise from natural variation or sampling.

To visualize the differences among the classes, radial visualization maps were plotted using the Radviz tool of Orange data mining toolbox [29]. Here we utilized the scores of select PCs obtained from subjecting the entire spectral dataset to PCA. Guided by the Vizrank algorithm, the PCs were chosen to maximize class separation. In the radial visualization plot, the scores of a spectrum determine the position of the corresponding data point relative to the PC pivots. Partial least squares discriminant analysis (PLS-DA), a supervised classification technique based on partial least squares regression, was employed to create decision models from the acquired Raman spectra for identifying the pre-metastatic niche [30]. PLS-DA-derived classification models were built and trained using a leave-m-out cross-validation approach that utilizes randomly chosen training data consisting of 60% of the data of each class and test data constituted by the remaining 40% of the spectra. Randomized equalization of classes was implemented prior to PLS-DA model development to avoid skewing the model through disproportionate class sizes. Multiple iterations of class equalization and splitting into testing and training sets (10x100) were performed to obtain average performance metrics of the PLS-DA derived classification models.

Collagen quantification of Masson's trichrome stained tissue slides was achieved using FIJI (Image-J based open source software) [31] and MATLAB (Mathworks, Natick, MA). The color de-convolution feature provided by FIJI was employed to extract an 8-bit frame (dense collagen presence = 0 and no collagen presence = 256) corresponding to the color, indicative of collagen content in the trichrome stains. The color was defined by average RGB values of pixels in a small user-selected region of interest (ROI) chosen in the image. Using in-house MATLAB code, the intensity of the pixels was converted to obtain a measure of collagen density in each frame. The data was averaged over the entire lung tissue section with n>35 fields of view (FOV) per class, where each FOV was *ca.* 1.75 mm x 1.33 mm. Statistical significance of differences across the classes was evaluated using the Student's t-test. A conventional criterion of p-value less than 0.05 was used to consider differences as statistically significant.

4.2.4 Microarray dataset

The gene expression microarray dataset GSE62817 from the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/) was used in this study [32]. This dataset contains gene expression data from pre-metastatic lungs of BALB/c mice injected with tumor cells into their fourth mammary fat pad. In particular, 67NR (non-metastatic) and 4T1 (metastatic) breast carcinoma cell lines were used and lung tissue was collected when the tumors reached a volume of 50 mm^3. Control mice with no tumor cell injections were utilized for comparison. Briefly, RNA was extracted using Qiagen kit, and Affymetrix microarrays (Mouse 430-v2) were used to analyze the expression profile of tissue samples. The heat map was generated using Gene-e matrix visualization and analysis software (http://www.broadinstitute.org). We used the moderated F-test statistic for selecting relevant genes. Consistent with the number of different groups and number of samples per group in the dataset, a threshold F-test statistic of 2.53 (corresponding to $\alpha = 0.125$ level of significance) was used.

4.3 Results and Discussion

Lung was selected as the target organ in the current pilot study, as it offers a favorable site for spontaneous dissemination of breast cancer and is the most commonly studied metastatic site in animal models [9, 33]. Primary orthotopic MDA-MB-231, and also eventually MCF-7, breast tumor xenografts used in our study preferentially metastasize to the lungs [34, 35]. Spectroscopic mapping of the lungs, as opposed to a limited number of discrete point measurements, was pursued to encompass a large field of view with high spectral contrast. This would also account spectroscopically for the intrinsic biological variation in lung tissue that could otherwise suppress the subtle differences expected from pre-metastatic adaptations. **Figure 4.1C** shows average Raman spectra recorded from lung samples of control mice ('Control') as well as mice bearing MCF-7 ('MCL') and MDA-MB-231 ('MDL') tumor xenografts. The spectra shown here were background subtracted for the tissue autofluorescence component. While gross visual inspection reveals limited spectral variations, we reason that a subset of pixels (representing specific molecular moieties) has predictive power that is lost in examining the average value of the spectra across the lung specimen. In an effort to focus on elucidating the differentiating biochemical characteristics, we employed principal component analysis (PCA). To preserve the subtle spectral features, we performed PCA on the normalized spectra recorded from the specimen without background subtraction. For comparison, the results obtained following fifth order best-fit polynomial based autofluorescence background removal have been also been provided alongside (and in Appendix).

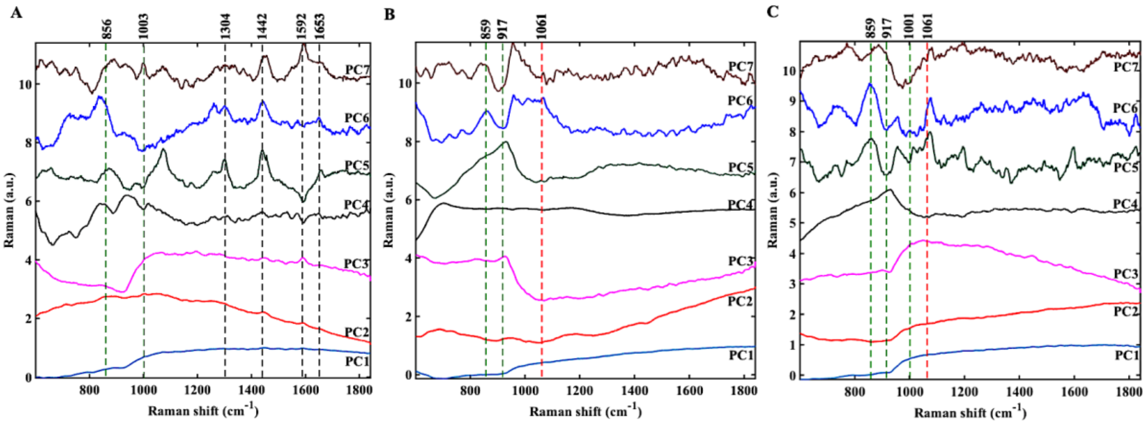


Figure 4.2: Principal component analysis of the acquired Raman spectra. (**A**) PC loadings derived from spectra of lungs from control mice, i.e. bearing no tumor xenograft. (**B**) PC loadings derived from spectra of lungs belonging to mice bearing MCF-7 xenografts (labeled as MCL in the text). (**C**) PC loadings derived from spectra of lungs belonging to mice with MDA-MB-231 xenografts (labeled as MDL in the text). Green and red dotted lines highlight collagen and proteoglycan features, respectively.

Consistent differences in Raman spectra reflect biochemical changes in pre-metastatic lungs. **Figure 4.2** shows the first 7 principal component (PC) loadings in order of spectral variance for each of the three classes, Control, MCL and MDL. The first few PCs in each class are evidently influenced by the broad tissue autofluorescence signal; the characteristic Raman features are more prevalent in PCs 4 through 7. The PCs derived from the spectra belonging to the lungs of control mice exhibit notable Raman features at 859 cm^{-1} (C-C stretch of proline in collagen), 1003 cm^{-1} (C–C stretching vibration of the aromatic ring in the phenylalanine side chain), 1442 cm^{-1} (CH_2 deformations in lipids), 1592 cm^{-1} (tentatively attributed to carbon particles) and 1653 cm^{-1} (amide-**I** feature of proteins with potential contributions of C=C stretching in lipids) with a weaker peak at 1304 cm^{-1} (in-plane CH_2 twisting modes of lipids). These features are concordant with prior observations in the literature [36-38]. **Table ST4.1** (Appendix) lists the prominent peaks observed in the PCs and their characteristic band assignments.

Visual inspection of the PC loadings show an enhancement of the 859 cm^{-1} peak, which can be attributed to collagen, for the MDL specimen in comparison with MCL and control as well as a new peak at 917 cm^{-1} (C-C stretch of proline ring) [17] for the non-control samples. These spectral differences suggest a positive correlation of collagen density in the lung specimens with the presence of a primary tumor xenograft and, importantly, with its metastatic potential. Previous studies have discussed the role of collagen in the pre-metastatic niche and have shown evidence of collagen crosslinking and the creation of a metastatic growth permissive fibrotic microenvironment at secondary sites, which was mediated by lysyl oxidase (LOX) secreted by hypoxic tumors [39, 40]. Inhibition of LOX synthesis in human breast cancer cells has been shown to reduce the accumulation of CD11b+ myeloid cells in pre-metastatic organs of mice with orthotopic tumors and prevent metastasis [12]. Another pertinent peak was observed at *ca.* 1061 cm^{-1} in the MCL and MDL PCs, which is known to be a key spectral marker for proteoglycans [41, 42]. This finding offers an intriguing insight into the nature of molecular modifications in the pre-metastatic niche, particularly in light of the study of Gao et al. This study demonstrated that myeloid cells in pre-metastatic lungs (recruited by primary tumor derived secretory factors) aberrantly expressed versican, an ECM proteoglycan [43]. Versican stimulated mesenchymal to epithelial transition of metastatic tumor cells by reducing phospho-Smad2 levels, which led to elevated cell proliferation and accelerated metastases. In fact, lung metastasis in mouse models was found to be significantly impaired through knockdown of versican, reinforcing the importance of proteoglycan content as a pre-metastatic site marker. Furthermore, the gradual increase in the prominence of proteoglycan marker in PCs with increasing metastatic potential is in agreement with the seminal report of Kaplan et al., which showed that recruitment of bone marrow-derived cells is correlated to the aggressiveness of the primary tumor [7]. On the other hand, a significant suppression of the peaks at *ca.* 1302 cm^{-1} and 1442 cm^{-1} was noted with a smaller reduction in the intensity of the 1653 cm^{-1} feature. Since the former two peaks are characteristic of lipids and the latter also has lipid contributions, one can reasonably infer a relative reduction in the lipid content corresponding to spectra from lungs of mice bearing primary tumor xenografts.

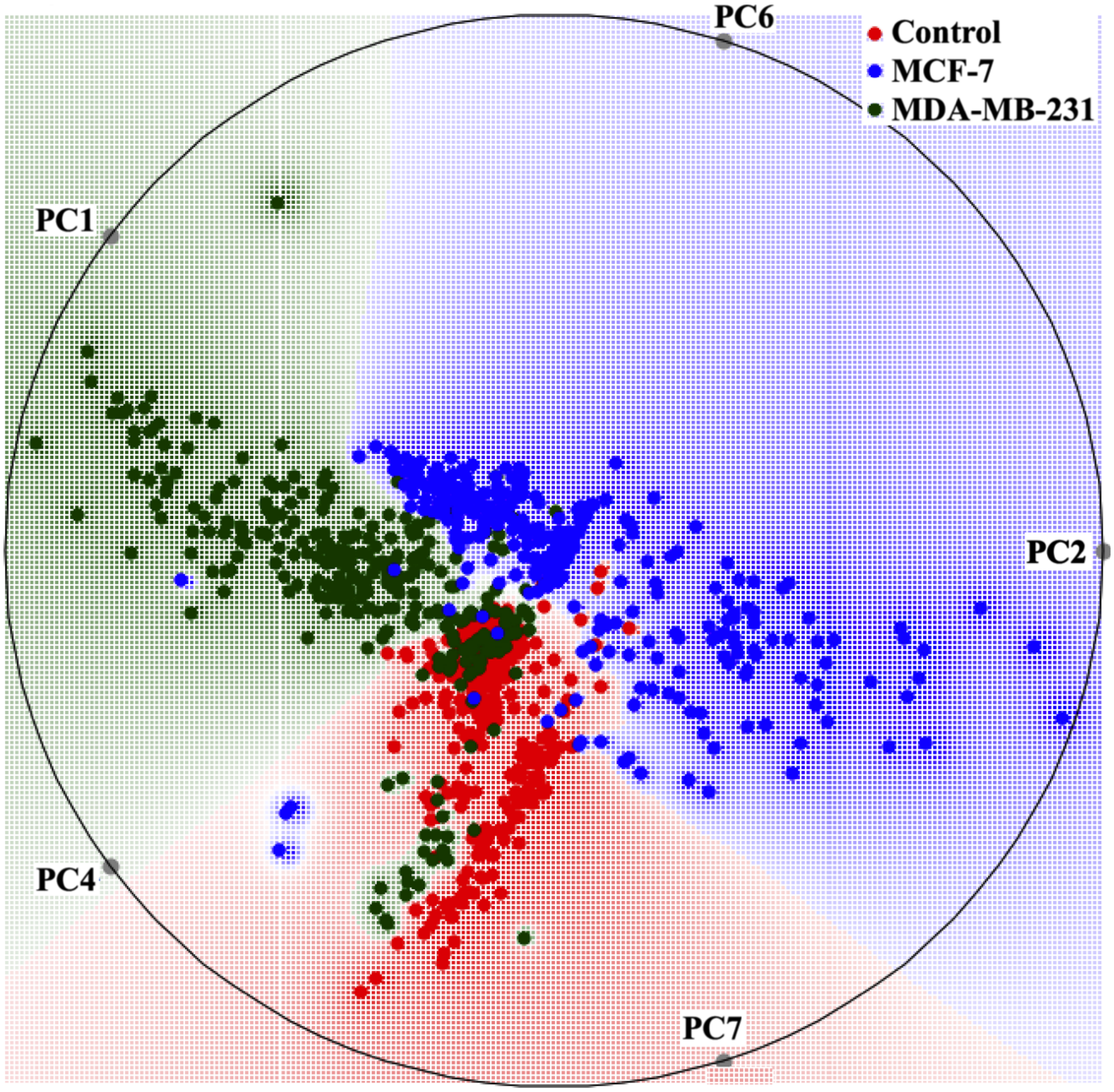


Figure 4.3: Visualization of spectroscopic differences due to pre-metastatic adaptations. Radial visualization plot showing clusters formed by spectra recorded from lung samples of sacrificed mice bearing MDA-MB-231 and MCF-7 breast cancer xenografts as well as controls without xenografts.

Given the large dimensionality of the spectral data, however, it is challenging to judge whether the differences across the classes are significant from visual inspection of the PC loadings alone. To observe these differences better, we employed radial visualization plots that map the scores of multiple PCs onto a two-dimensional space for the purpose of clustering. **Figure 4.3** shows a representative radial visualization plot constructed by using PCs derived from a randomized spectral selection with 300 points per class (control, MCL and MDL). These were chosen from the total set consisting of *ca.* 900 spectra/class, which in turn were constituted by *ca.* 300 spectra acquired from spatially distinct points in the lung lobes of each mouse. **Figure S4.1** (Appendix) shows the corresponding radial visualization map after subtraction of tissue autofluorescence background.

In order to obtain informative projections of the class-labeled data, the VizRank algorithm was used to grade the PCs by their ability to visually discriminate between classes [44]. Evidently, there are pronounced differences in the Raman spectra acquired from lung specimens of control, MCL and MDL mouse models, most likely owing to differential priming through factors secreted by the primary tumor. The presence of a small overlap of clusters from control and MDL mice indicates a limited development of the pre-metastatic niche in some of the latter cases and requires further analysis, as detailed in the ensuing paragraphs. While the PC score-based plot offers a satisfactory tool for preliminary data exploration, it does not provide quantitative information about the potential of Raman spectroscopy in recognizing the class (metastatic potential) and in understanding how the lung(s) of an individual mouse responds to the primary tumor xenograft.

Thus, we used partial least squares–discriminant analysis (PLS-DA)-based classification models for translating the spectroscopic measurements in the pre-metastatic lungs to identification of the type of primary tumor xenograft. We employed an equal number of spectra belonging to each class (control, MCL and MDL) and their class labels to train the classification algorithm. To ensure robustness, we evaluated the classifier by testing on a separate validation dataset as detailed in the Data Analysis section. Average correct rates of prediction of 90.1%, 97.7% and 78.4% (95.4%, 95.6% and 75.1% after autofluorescence background subtraction) were obtained for the spectra belonging to control, MCL and MDL, respectively. The relevant confusion matrix of the reference and predicted labels is shown in **Table ST4.2** (Appendix). The lower correct classification rate for MDL spectra in both the cases is in agreement with the overlap of the MDL and control clusters observed on the radial visualization plot in **Fig. 4.3**.

In order to understand the root cause of the MDL spectra misclassifications, we repeated the former analysis by leaving one mouse out of the dataset each time (**Table 4.1** and **Table ST4.3** in Appendix after autofluorescence background subtraction). Removing mouse MD #3 (arbitrary numbering of mice used for tabulating results) yields near perfect classification accuracy indicating significantly lesser pre-metastatic adaptations in the lungs of this animal. Furthermore, removing mouse MD #3 also improved the classification rate of spectra belonging to control mice due to enhanced contrast in the training data. Notably, removal of any other mouse from the classification protocol did not result in as significant a change in the accuracy levels. This reinforces the fact that the improvement observed on removal of mouse MD #3 data was not due to overtraining of the model on smaller numbers, as otherwise similar enhancements would have been noted in all the other cases. The inadequate priming of the MD #3 lungs is also supported by application of Chauvenet's criterion to the set of classification rates obtained for the MDL class (**Table 4.1**). The latter results in designation of MD #3 as the sole outlier in the group due to its significant deviation from the mean by more than the maximum allowable number of standard deviations (τ_{max} = 1.96 for a sample size of n = 10). Application of Chauvenet's criterion also facilitates determination of individual sample eligibility for training the PLS-DA classifier. The spectroscopic measurements, thus, capture the inherent variability in metastasis, which is commonly regarded as an inefficient process that only a subset of tumor cells can successfully navigate [45, 46] and is known to exhibit sporadic occurrence across a cohort of animals.

Finally, we conducted a negative control study to verify that the predictive power of the developed algorithms was not driven by potential spurious correlations in the spectral dataset [47]. For this validation study, we assigned random class labels to the spectra irrespective of their true class origins and employed the PLS-DA-derived classification models after similar splitting of the data into training and test sets. This resulted in an average correct classification rate of 33.3% with a standard deviation of 1.4% (and 33.6% with a standard deviation of 1.4% after background subtraction) for 1000 iterations. The significantly low rate of correct classification (consistent with the likelihood of random selection of the true class label, 1/3) underscores the absence of chance correlations in the developed model.

Table 4.1: Correct classification rates (%) of the PLS-DA-derived model using leave-one-mouse-out protocol (MD and MC refer to mouse models with MDA-MB-231 and MCF-7 tumor xenografts, respectively)

Mouse excluded	**Correct classification rate (%)**			**Chauvenet's criterion for MDL** (n = 10; τ_{max} = 1.96)	
	Control	**MCL**	**MDL**	$\tau = \lvert x_i - x_{mean} \rvert / \sigma$	**Result**
None	90.1	97.7	78.4	0.26	Retain
MD #1	81.2	97.0	75.7	0.61	Retain
MD #2	83.8	96.8	76.2	0.54	Retain
MD #3	100.0	98.6	99.4	2.54	Eliminate
MC #1	88.8	98.0	78.5	0.24	Retain
MC #2	89.6	97.3	77.6	0.36	Retain
MC #3	88.8	98.3	77.3	0.40	Retain
Control #1	87.2	96.4	73.1	0.96	Retain
Control #2	92.9	97.3	80.2	0.01	Retain
Control #3	92.9	97.4	86.4	0.81	Retain

Histological assessment of the pre-metastatic niche in mice lungs. Due to their high metastatic potential and preference for metastasis to lungs, orthotopic MDA-MB-231 xenografts are frequently employed to replicate breast cancer metastasis and organotropism [33, 48]. Aggressive subpopulations of MDA-MB-231 are often derived through multiple rounds of in vivo selection and re-implantation and have been recently reported to result in macro-metastases to the lungs in 100% of all tested mice [35]. In our study, we observed no cancer cell seeding in lungs of mice bearing MDA-MB-231 tumor xenografts (time of sacrifice: 8-12 weeks post orthotopic tumor inoculation). Prior optical tracking studies by Winnard and co-workers showed that orthotopically implanted MDA-MB-231 cells reached lungs only after ~15 weeks of implantation in SCID mice [34]. They also observed the absence of distant metastases after 8 weeks, consistent with the time period of sacrifice in our study. MCF-7 cells, often classified as non-metastatic [49], were likewise not expected to engraft in the lungs within this 8-12 week time frame. However, it is noteworthy that MCF-7 cells are known to eventually metastasize to lungs in immunodeficient mice such as NSG [35].

Here, the lung tissue sections from each mouse were H&E stained to check for the onset (or the lack thereof) of metastasis. Also, to histologically examine the differences in collagen content across the classes, serial sections were processed with Masson’s trichrome stain. **Figure 4.4** shows representative images of H&E and Masson’s trichrome stained lung sections belonging to each class (control, MCL and MDL). The H&E images corroborate the lack of any metastatic lesions in the lung specimens.

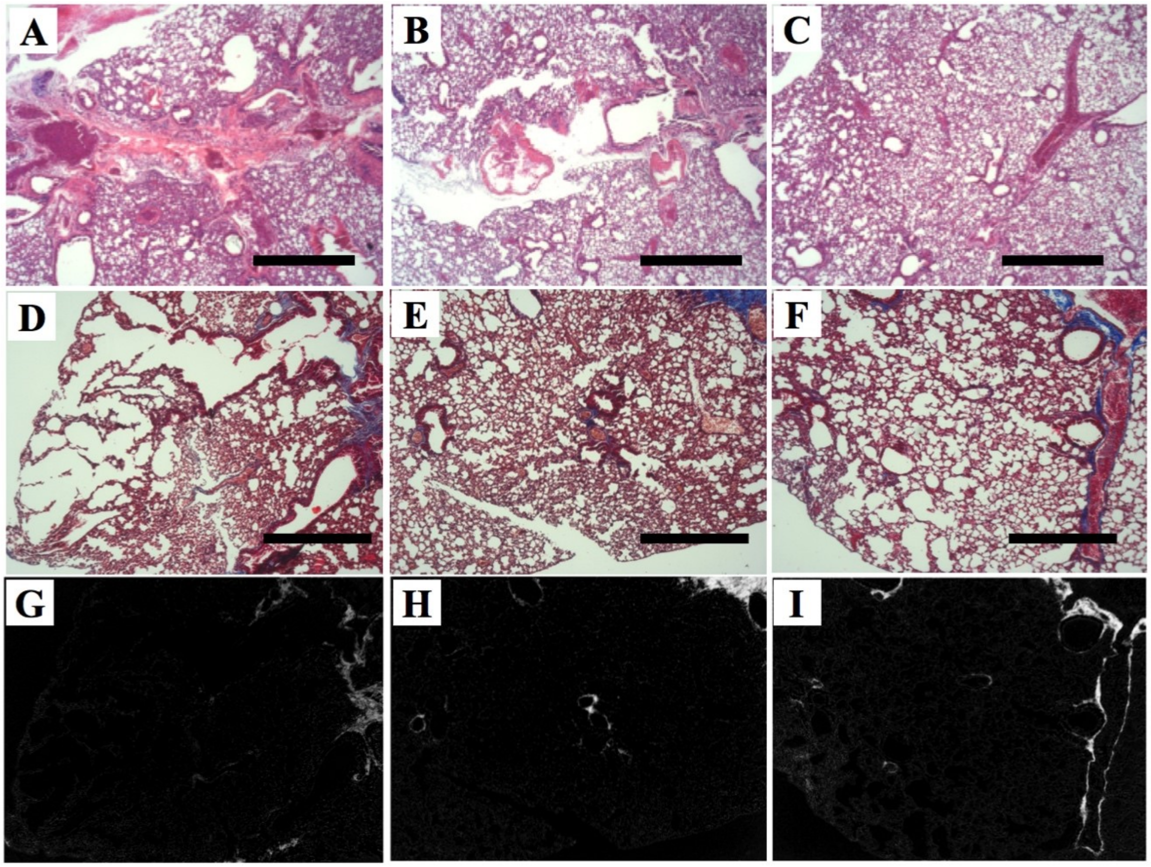


Figure 4.4: Histological assessment of pre-metastatic lungs shows stromal changes. Top (**A-C**) and middle (**D-F**) panels display representative microscopic images of H&E and Masson's trichrome stained slides at 5x and 10x magnifications, respectively. The H&E stained sections confirm the absence of tumor cell seeding in the lungs of controls. Masson's trichrome stain delineates collagen fibers in the extracellular matrix and is quantified through image processing, as shown in the bottom panel (**G-I**). The left (**A, D, G**) panel shows lung sections derived from control mice whereas the middle (**B, E, H**) and right (**C, F, I**) panels represent lung sections from mice bearing MCF-7 (non-metastatic) and MDA-MB-231 (metastatic) tumor xenografts, respectively. The scale bars in the top and middle panels represent 1,000 and 500 µm, respectively.

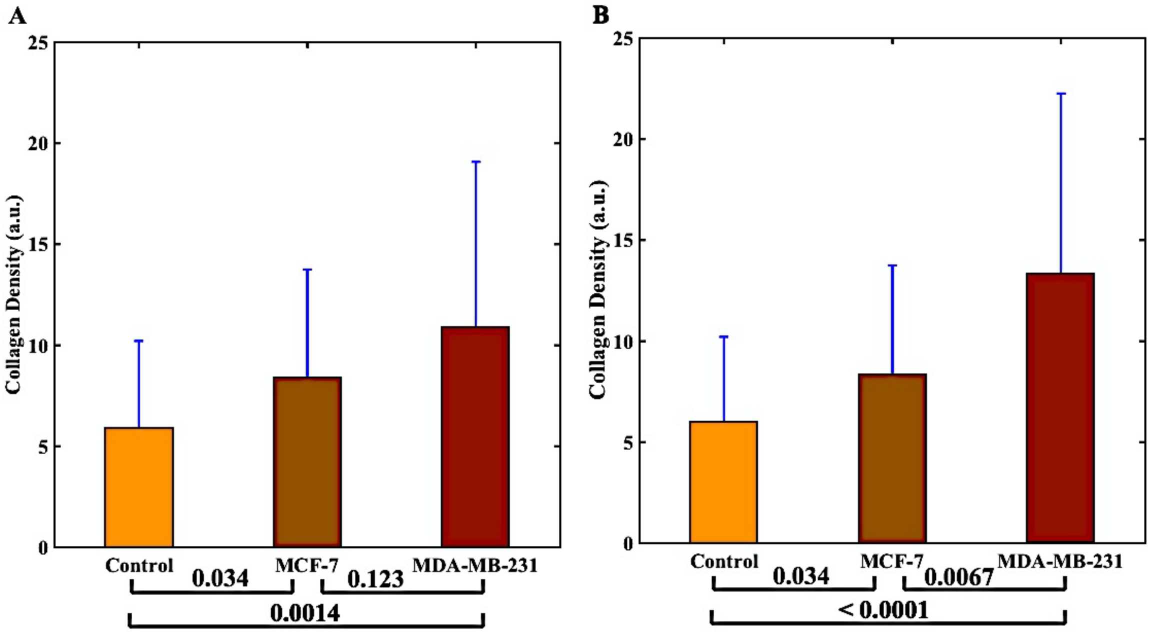
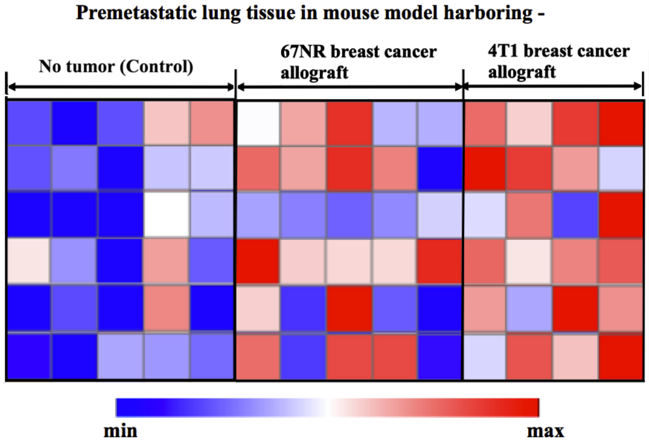


Figure 4.5: Quantification of collagen fiber density in pre-metastatic lungs. (**A**) Bar plot showing mean and standard deviation of collagen density across the three classes (with all mice included) along with pairwise Student's t-test p-values. (**B**) Bar plot showing mean and standard deviation of collagen content across the three classes (after exclusion of MDA-MB-231 xenograft bearing mouse displaying atypical Raman data) along with pairwise Student's t-test p-values.

The Masson's trichrome stained sections were used for quantification of the mean collagen density for each class (detailed in Materials and Methods). **Figure 4.5A** shows the mean bar plot that highlights the differences in collagen density for control mice and mice bearing MCF-7 and MDA-MB-231 xenografts. We observe that the metastatic potential of the primary tumor is positively correlated with the collagen density in the pre-metastatic niche. Yet, the differences in the mean collagen density values between MCL and MDL samples do not reach statistical significance ($p<0.05$). Based on our spectroscopic findings, we suspected that the lung specimens of mouse MD #3 may possibly skew the collagen density values of the MDL set. Accordingly, we re-calculated the values by removing the images of the lungs of this mouse, as shown in **Fig. 4.5B**. With this modification, the differences among each pair of classes were found to be statistically significant. This improvement of contrast in collagen density corresponds well with our spectroscopic findings and reflects the biochemical sensitivity of the vibrational spectroscopic data.

In light of the spectroscopic identification of stromal adaptations, we further sought to investigate the genetic underpinnings of pre-metastatic priming of lungs. We performed gene expression analysis on publicly available microarray data (GSE 62817) to determine markers in pre-metastatic lungs in response to primary breast tumors of divergent metastatic potential [32]. Specifically, the data included gene expression levels corresponding to the lungs of normal mice (n=5) as well as pre-metastatic lungs of mice injected with non-metastatic 67NR breast carcinoma cells (n=5) and with metastatic 4T1 breast carcinoma cells (n=4). Seeking to isolate genes relevant to our study, we restricted our search to genes encoding for key stromal constituents and significantly overexpressed in pre-metastatic lungs of 4T1 tumor bearing mice. **Figure 4.6** shows the heatmap representing expression levels of these genes along with corresponding moderated F-statistic. Pre-metastatic lungs of the 4T1 tumor bearing mice demonstrate a selective upregulation of genes related to ECM constituents, notably collagen, fibronectin, versican and glypican. Importantly, each of these ECM components exhibits a decreasing gradient of values from 4T1 to 67NR and then to control cases. The differential expression of stromal genes in response to primary tumor development can, thus, help explain our observations of discernible biochemical alterations in pre-metastatic lungs of mice bearing MCF-7 xenografts, even though these cells rarely metastasize in the mouse model used.



Symbol	Description	ID	Mod. F-Value
Col15a1	Collagen, type XV, alpha 1	1448755_at	3.88
Col5a1	Collagen, type V, alpha 1	1416741_at	3.83
Col4a3	Collagen, type IV, alpha 3	1438779_at	2.55
Fn1	Fibronectin 1	1426642_at	3.31
Vcan	Versican	1421694_a_at	3.61
Gpc4	Glypican 4	1443620_at	4.25

Figure 4.6. Gene expression changes in pre-metastatic lungs as a function of metastatic potential of primary tumor. Microarray gene expression data heat map was obtained by analyzing the publicly available dataset GSE62816 on the Gene-e data visualization and analysis platform. The sample cohort includes lungs of mice bearing breast tumor xenografts of different metastatic potential. Total RNA was isolated from the pre-metastatic lungs and hybridized on an Affymetrix Mouse Genome 430 2.0 Array. Genes that are relevant to spectral markers identified in the current study and overexpressed in response to the metastatic potential of the primary tumor were analyzed. Moderated F-value of 2.53 was set as the criterion for inclusion.

Taken together, our findings suggest that remodeling of the ECM such as an increase in collagen and proteoglycan content occurs in response to primary tumor derived factors, which precedes the actual seeding of tumor cells at the distant metastatic site. The data in this study support a continuous pre-metastatic niche formation model from primary tumors with low and high metastatic potential, rather than discrete pre-metastatic adaptations that are representative of the highly metastatic model alone. This would also imply that pre-metastatic adaptations are a necessary condition for further progression but not predictive of the eventual success of metastases.

In conclusion, the current study proposes Raman spectroscopy as a label-free molecular-specific tool for detection of pre-metastatic adaptations in the stromal environment. Using breast cancer metastasis to the lungs as the paradigm, we have demonstrated that Raman spectroscopy accurately detects changes in the ECM of pre-metastatic lungs, which correlate with the metastatic potential of the respective primary tumor xenograft. We identified spectral markers corresponding to collagen and proteoglycan that offer molecular insights into the formation of the pre-metastatic niche while also facilitating objective detection. The data presented here are unique and complementary to other microenvironment profiling methods such as genomic assays and mass spectrometry. While breast cancer metastasis to the lungs has been chosen for the current study, it should be noted that this approach can be extended to study the development of pre-metastatic niches at any secondary target organ from primary breast and non-breast malignancies.

We envision that the use of Raman spectroscopic imaging in conjunction with further biochemical assays will offer detailed mechanistic insights into pre-metastatic niche formation and evolution. As such, this offers a unique research tool that combines microenvironment and cellular profiling through non-perturbative, multiplexed measurements of proteins, nucleic acids, lipids and metabolites. Building on the ability to detect such subtle changes in tissue composition, and as discussed in recent reports [18, 23], we anticipate that Raman spectroscopic imaging can, with further refinement, facilitate surgical margin assessment in tissue conserving surgery and provide prediction of tumor recurrence. Integration of Raman spectroscopy with minimally invasive biopsy needles can also permit real-time, in situ detection of malignancies [19, 50].

4.4 Appendix

Table ST4.1: Band assignment for spectral features observed in PC loadings derived from the Raman spectra of the mouse lungs

Observed Raman Peaks in the PC loadings (cm^{-1})	Raman band assignment from literature
859	C-C stretch of proline (Collagen)
917	C-C stretch of proline ring (Collagen)
1003	C–C stretching vibration of the aromatic ring in the phenylalanine side chain (Collagen)
1061	OSO_3 symmetric stretching (Proteoglycans)
1304	In-plane CH_2 twisting modes of lipids
1442	CH_2 deformation of lipids
1592	Carbon particles (source of particles unknown)
1653	Amide I (symmetric C=O stretching mode of proteins) and C=C lipid stretch (Lipids and proteins)

Table ST4.2: Confusion matrix for PLS-DA derived classification model showing correct classification rates (%, averaged over 1000 iterations)

		Predicted Class Labels		
		Control	**MCL**	**MDL**
Reference Labels	**Control**	90.1	0.1	9.8
	MCL	1.0	97.7	1.3
	MDL	20.9	0.7	78.4

Table ST4.3: Correct classification rates (%) of the PLS-DA-derived model using leave-one-mouse-out protocol with background subtracted spectra (MD and MC refer to mouse models with MDA-MB-231 and MCF-7 tumor xenografts, respectively)

Mouse excluded	**Correct classification rate (%)**			**Chauvenet's criterion for MDL** (n = 10; τ_{max} = 1.96)	
	Control	**MCL**	**MDL**	$\tau = \lvert x_i - x_{mean} \rvert / \sigma$	**Result**
None	95.4	95.6	75.1	0.71	Retain
MD #1	85.7	94.2	81.5	0.04	Retain
MD #2	90.1	94.8	76.8	0.52	Retain
MD #3	100.0	96.0	100.0	2.18	Eliminate
MC #1	88.5	96.3	83.0	0.21	Retain
MC #2	96.3	95.5	74.5	0.77	Retain
MC #3	96.2	96.5	73.1	0.95	Retain
Control #1	96.7	93.4	73.7	0.87	Retain
Control #2	98.0	94.0	84.8	0.42	Retain
Control #3	95.6	94.1	89.6	0.97	Retain

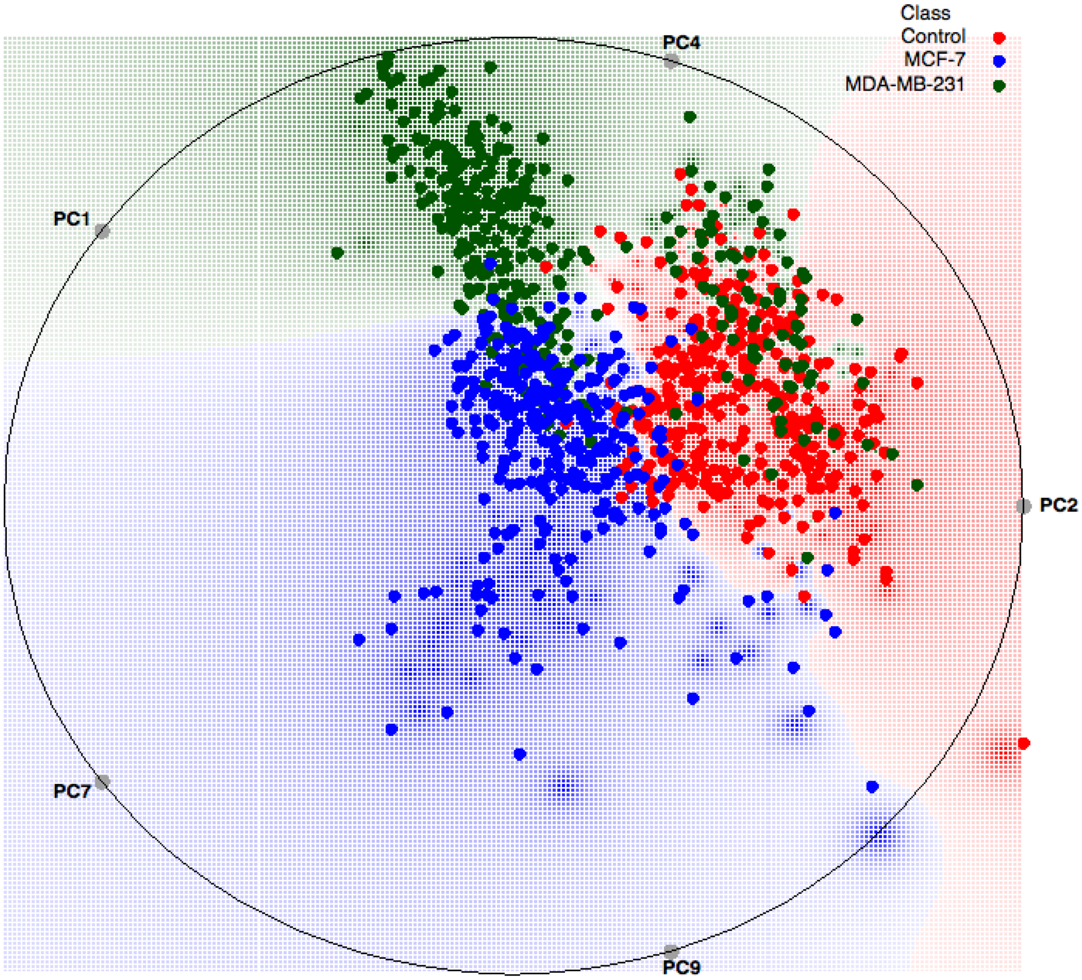


Figure S4.1: Radial visualization plot after background subtraction. Radial visualization plot showing clusters formed by spectra recorded from lung samples of sacrificed mice bearing MDA-MB-231 and MCF-7 breast cancer xenografts as well as controls without xenografts after autofluorescence background subtraction.

Bibliography

[1] M. B. Sporn, The war on cancer, *The Lancet* **347**(9012), 1377-1381.

[2] K. R. Hess, G. R. Varadhachary, S. H. Taylor, W. Wei, M. N. Raber, R. Lenzi, et al., Metastatic patterns in adenocarcinoma, *Cancer* **106**(7), 2006, 1624-1633.

[3] Y. N. M. Lee, Breast carcinoma: Pattern of metastasis at autopsy, *J Surg Oncol* **23**(3), 1983, 175-180.

[4] S. Paget, The distribution of secondary growths in cancer of the breast. 1889, *Cancer Metastasis Rev* **8**(2), 1989, 98-101.

[5] I. J. Fidler and M. L. Kripke, Metastasis results from preexisting variant cells within a malignant tumor, *Science* **197**(4306), 1977, 893-895.

[6] I. J. Fidler, Selection of successive tumour lines for metastasis, *Nature* **242**(118), 1973, 148-149.

[7] R. N. Kaplan, R. D. Riba, S. Zacharoulis, A. H. Bramley, L. Vincent, C. Costa, et al., VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche, *Nature* **438**(7069), 2005, 820-827.

[8] S. Hiratsuka, A. Watanabe, H. Aburatani and Y. Maru, Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis, *Nat Cell Biol* **8**(12), 2006, 1369-1375.

[9] B. Psaila and D. Lyden, The metastatic niche: Adapting the foreign soil, *Nat Rev Cancer* **9**(4), 2009, 285-293.

[10] B. Costa-Silva, N. M. Aiello, A. J. Ocean, S. Singh, H. Zhang, B. K. Thakur, et al., Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver, *Nat Cell Biol* **17**(6), 2015, 816-826.

[11] S. Hiratsuka, K. Nakamura, S. Iwai, M. Murakami, T. Itoh, H. Kijima, et al., MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis, *Cancer Cell* **2**(4), 2002, 289-300.

[12] J. T. Erler, K. L. Bennewith, T. R. Cox, G. Lang, D. Bird, A. Koong, et al., Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche, *Cancer Cell* **15**(1), 35-44.

[13] J. T. Erler, K. L. Bennewith, M. Nicolau, N. Dornhöfer, C. Kong, Q. Le, et al., Lysyl oxidase is essential for hypoxia-induced metastasis, *Nature* **440**(7088), 2006, 1222-1226.

[14] P. Matousek and N. Stone, Development of deep subsurface raman spectroscopy for medical diagnosis and disease monitoring, *Chem Soc Rev* **45**(7), 2016, 1794-1802.

[15] D. C. Fernandez, R. Bhargava, S. M. Hewitt and I. W. Levin, Infrared spectroscopic imaging for histopathologic recognition, *Nat Biotechnol* **23**(4), 2005, 469-474.

[16] M. J. Walsh, R. K. Reddy and R. Bhargava, Label-free biomedical imaging with mid-ir spectroscopy, *Selected Topics in Quantum Electronics, IEEE Journal Of* **18**(4), 2012, 1502-1513.

[17] N. Stone, C. Kendall, N. Shepherd, P. Crow and H. Barr, Near-infrared raman spectroscopy for the classification of epithelial pre-cancers and cancers, *J Raman Spectrosc* **33**(7), 2002, 564-573.

[18] K. Kong, C. J. Rowlands, S. Varma, W. Perkins, I. H. Leach, A. A. Koloydenko, et al., Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and raman scattering microscopy, *Proc Natl Acad Sci U S A* **110**(38), 2013, 15189-15194.

[19] M. Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, et al., Intraoperative brain cancer detection with raman spectroscopy in humans, *Sci Transl Med* **7**(274), 2015, 274ra19.

[20] C. Matthäus, C. Krafft, B. Dietzek, B. R. Brehm, S. Lorkowski and J. Popp, Noninvasive imaging of intracellular lipid metabolism in macrophages by raman microscopy in combination with stable isotopic labeling, *Anal Chem* **84**(20), 2012, 8549-8556.

[21] R. Pandey, S. K. Paidi, J. W. Kang, N. Spegazzini, R. R. Dasari, T. A. Valdez, et al., Discerning the differential molecular pathology of proliferative middle ear lesions using raman spectroscopy, *Scientific Reports* **5**2015, 13305.

[22] D. C. Malins, K. M. Anderson, N. K. Gilman, V. M. Green, E. A. Barker and K. E. Hellström, Development of a cancer DNA phenotype prior to tumor formation, *Proceedings of the National Academy of Sciences of the United States of America* **101**(29), 2004, 10721-10725.

[23] J. T. Kwak, A. Kajdacsy-Balla, V. Macias, M. Walsh, S. Sinha and R. Bhargava, Improving prediction of prostate cancer recurrence using chemical imaging, *Scientific Reports* **5**2015,

[24] A. Rizwan, C. Bulte, A. Kalaichelvan, M. Cheng, B. Krishnamachary, Z. M. Bhujwalla, et al., Metastatic breast cancer cells in lymph nodes increase nodal collagen density, *Scientific Reports* **5**2015,

[25] A. Rizwan, I. Serganova, R. Khanin, H. Karabeber, X. Ni, S. Thakur, et al., Relationships between LDH-A, lactate, and metastases in 4T1 breast tumors, *Clin Cancer Res* **19**(18), 2013, 5158-5169.

[26] R. I. Yarden, A. H. Lauber, D. El-Ashry and S. A. Chrysogelos, Bimodal regulation of epidermal growth factor receptor by estrogen in breast cancer cells, *Endocrinology* **137**(7), 1996, 2739-2747.

[27] S. K. Paidi, S. Siddhanta, R. Strouse, J. B. McGivney, C. Larkin and I. Barman, Rapid identification of biotherapeutics with label-free raman spectroscopy, *Anal Chem* **88**(8), 2016, 4361-4368.

[28] I. Jolliffe, Principal component analysis, Wiley Online Library, 2002,

[29] J. Demšar, T. Curk, A. Erjavec, Č Gorup, T. Hočevar, M. Milutinovič, et al., Orange: Data mining toolbox in python, *The Journal of Machine Learning Research* **14**(1), 2013, 2349-2353.

[30] R. G. Brereton, Applied chemometrics for scientists, John Wiley & Sons, 2007,

[31] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, et al., Fiji: An open-source platform for biological-image analysis, *Nature Methods* **9**(7), 2012, 676-682.

[32] M. Kowanetz, X. Wu, J. Lee, M. Tan, T. Hagenbeek, X. Qu, et al., Granulocyte-colony stimulating factor promotes lung metastasis through mobilization of Ly6G+Ly6C+ granulocytes, *Proc Natl Acad Sci U S A* **107**(50), 2010, 21248-21255.

[33] A. J. Minn, G. P. Gupta, P. M. Siegel, P. D. Bos, W. Shu, D. D. Giri, et al., Genes that mediate breast cancer metastasis to lung, *Nature* **436**(7050), 2005, 518-524.

[34] P. T. Winnard, J. B. Kluth and V. Raman, Noninvasive optical tracking of red fluorescent protein-expressing cancer cells in a model of metastatic breast cancer, *Neoplasia* **8**(10), 2006, 796-IN1.

[35] E. Iorns, K. Drews-Elger, T. M. Ward, S. Dean, J. Clarke, D. Berry, et al., A new mouse model for the study of human breast cancer metastasis, *PLoS One* **7**(10), 2012, e47995.

[36] N. Huang, M. Short, J. Zhao, H. Wang, H. Lui, M. Korbelik, et al., Full range characterization of the raman spectra of organs in a murine model, *Optics Express* **19**(23), 2011, 22892-22909.

[37] B. G. Frushour and J. L. Koenig, Raman scattering of collagen, gelatin, and elastin, *Biopolymers* **14**(2), 1975, 379-391.

[38] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor and M. Baranska, Raman spectroscopy of lipids: A review, *J Raman Spectrosc* **46**(1), 2015, 4-20.

[39] Y. N. Joo, H. Jin, S. Y. Eun, S. W. Park, K. C. Chang and H. J. Kim, P2Y2R activation by nucleotides released from the highly metastatic breast cancer cell MDA-MB-231 contributes to pre-metastatic niche formation by mediating lysyl oxidase secretion, collagen crosslinking, and monocyte recruitment, *Oncotarget* **5**(19), 2014, 9322-9334.

[40] T. R. Cox, D. Bird, A. M. Baker, H. E. Barker, M. W. Ho, G. Lang, et al., LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis, *Cancer Res* **73**(6), 2013, 1721-1732.

[41] A. Kunstar, J. Leijten, S. van Leuveren, J. Hilderink, C. Otto, C. A. van Blitterswijk, et al., Recognizing different tissues in human fetal femur cartilage by label-free raman microspectroscopy, *J Biomed Opt* **17**(11), 2012, 116012.

[42] S. Gamsjaeger, K. Klaushofer and E. P. Paschalis, Raman analysis of proteoglycans simultaneously in bone and cartilage, *J Raman Spectrosc* **45**(9), 2014, 794-800.

[43] D. Gao, N. Joshi, H. Choi, S. Ryu, M. Hahn, R. Catena, et al., Myeloid progenitor cells in the premetastatic lung promote metastases by inducing mesenchymal to epithelial transition, *Cancer Res* **72**(6), 2012, 1384-1394.

[44] G. Leban, B. Zupan, G. Vidmar and I. Bratko, Vizrank: Data visualization guided by machine learning, *Data Mining and Knowledge Discovery* **13**(2), 2006, 119-136.

[45] L. Weiss, Metastatic inefficiency, *Adv Cancer Res* **54**(159211.7), 1990,

[46] C. W. Wong, A. Lee, L. Shientag, J. Yu, Y. Dong, G. Kao, et al., Apoptosis: An early event in metastatic inefficiency, *Cancer Res* **61**(1), 2001, 333-338.

[47] M. A. Arnold, J. J. Burmeister and G. W. Small, Phantom glucose calibration models from simulated noninvasive human near-infrared spectra, *Anal Chem* **70**(9), 1998, 1773-1781.

[48] X. Lu and Y. Kang, Efficient acquisition of dual metastasis organotropism to bone and lung through stable spontaneous fusion between MDA-MB-231 variants, *Proceedings of the National Academy of Sciences* **106**(23), 2009, 9385-9390.

[49] A. Rizwan, M. Cheng, Z. M. Bhujwalla, B. Krishnamachary, L. Jiang and K. Glunde, Breast cancer cell adhesome and degradome interact to drive metastasis, *NPJ Breast Cancer* **1**2015, 15017.

[50] I. Barman, N. C. Dingari, A. Saha, S. McGee, L. H. Galindo, W. Liu, et al., Application of raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy, *Cancer Research* **73**(11), 2013, 3206-3215.

Chapter 5

Label-free Raman spectroscopy reveals signatures of radiation resistance in the tumor microenvironment

Delay in the assessment of tumor response to radiation therapy continues to pose a major challenge to quality of life for patients with non-responsive tumors. Here we exploited label-free Raman spectroscopic mapping to elucidate radiation-induced biomolecular changes in tumors and uncovered latent microenvironmental differences between treatment-resistant and -sensitive tumors. We used isogenic radiation-resistant and -sensitive A549 human lung cancer cells human head and neck squamous cell carcinoma (HNSCC) cell lines (UM-SCC-47 and UM-SCC-22B, respectively) to grow tumor xenografts in athymic nude mice and demonstrated the molecular specificity and quantitative nature of Raman spectroscopic tissue assessments. Raman spectra obtained from untreated and treated tumors were subjected to chemometric analysis using multivariate curve resolution-alternating least squares (MCR-ALS) and support vector machine (SVM) to quantify biomolecular differences in the tumor microenvironment. The Raman measurements revealed significant and reliable differences in lipid and collagen content post-radiation in the tumor microenvironment, with consistently greater changes observed in the radiation-sensitive tumors. In addition to accurately

---

This chapter is a reprint of the peer-reviewed publication: Paidi SK*, Diaz PM*, Dadgar S*, Jenkins SV, Quick CM, Griffin RJ, Dings RPM, Rajaram N, Barman I. “Label-free Raman spectroscopy reveals tumor microenvironmental signatures of radiation resistance” Cancer Research, 79(8), 2054-64, 2019. (* denotes equal contribution) evaluating tumor response to therapy, the combination of Raman spectral markers potentially offers a route to predicting response in untreated tumors prior to commencing treatment. Combined with its non-invasive nature, our findings provide a rationale for in vivo studies using Raman spectroscopy, with the ultimate goal of clinical translation for patient stratification and guiding adaptation of radiotherapy during the course of treatment.

5.1 Introduction

Radiation in conjunction with chemotherapy or other targeted therapies is used to treat the majority of lung and head and neck cancer patients. The overall radiation dose is fractionated and delivered over a period of 5-7 weeks (2 Gy/day, 5 days/week) because dose fractionation is believed to improve tumor oxygenation and, hence, overall cell kill [1, 2]. An outstanding challenge in optimizing the efficacy of such treatment resides in determining the degree of radiosensitivity associated with a specific patient's disease and the extent of tumor response to radiation. There are no accepted methods to determine treatment response either before or during the early stages of therapy. Although Human Papilloma Virus (HPV)-negative head and neck squamous cell carcinomas (HNSCCs) are associated with significantly worse outcomes compared with HPV-positive tumors [3, 4], HPV status is not used to guide treatment of HNSCC. Currently, X-ray Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) is used to determine tumor shrinkage about 2-3 weeks after completion of therapy. Positron Emission Tomography (PET) of fluorodeoxyglucose (FDG) uptake to measure functional tumor response is recommended about 8-12 weeks after completion of therapy to avoid false positives. Hence, patients who undergo the full treatment regimen and are later identified as non-responders are exposed to the toxic side effects of ineffective therapy for the full duration of the treatment regimen. Identifying patients with radiation-resistant tumors, prior to commencing treatment or immediately after, would significantly improve treatment response rates and help non-responding patients avoid the toxic side effects of ineffective radiation therapy.

Seeking to address this unmet need, molecular alterations in the tumor microenvironment in response to radiation therapy have been studied from multiple points of view including tumor hypoxia [5-7], cell repopulation [8-10], and genetic mutations involved in DNA repair pathways [11]. However, elucidation of serum and/or imaging biomarkers for accurate patient stratification and continuous assessment of therapy response, and their translation to the clinic has proven to be challenging. In an effort to develop better phenotypic strategies that could aid the clinical practice of radiation oncology, we propose an entirely complementary optical tool to the existing imaging arsenal featuring Raman scattering to non-invasively quantify the putative differences in the molecular milieu of radiosensitive and radioresistant tumors.

Raman spectroscopy offers a non-ionizing, label-free and highly specific technique for molecular characterization of the tumor and its microenvironment [12, 13]. It relies on the inelastic scattering of light, arising from its interactions with the biological specimen, to quantify the unique vibrational modes of molecules within its native context [14]. Raman spectroscopy offers the ability to probe biomolecular changes both in vivo and ex vivo, and interrogate complex molecular heterogeneity directly from cells and tissues [15]. Recent studies by us and others have harnessed vibrational profiles for objective recognition of epithelial and stromal changes in cancers [16-22]. Emerging data suggests the presence of post-radiation alterations in Raman spectral features and biomolecular differences between cell lines of varying radiosensitivity [23, 24]. Krishna and colleagues showed that radiation-induced changes in Raman spectra could be used to differentiate treatment responders and non-responders in excised cervical cancers; however, pre-treatment Raman spectra were incapable of identifying radiation response [25]. Furthermore, a recent Raman spectroscopic study on ex vivo tumor xenografts by Jirasek and co-workers identified elevated levels of glycogen in tumors exposed to a single, high radiation dose of 15 Gy [26]. While these reports underscore the promise of Raman spectroscopy in detecting radiation-induced changes, these measurements were performed on cells or tumor xenografts following a single radiation dose. More systematic studies that examine the sensitivity of Raman spectroscopy to changes in the tumor microenvironment when subjected to fractionated, clinically relevant radiation doses have been lacking. Such measurements would provide a better understanding of molecular modifications resulting from fractionated dosing and, ultimately, facilitate a personalized treatment approach. In addition, spectral markers of intrinsic radiation resistance that can be identified in tumors even before commencing therapy could provide a paradigm shift in determining treatment regimen.

The goal of our study was to leverage Raman spectroscopy to investigate biomolecular changes within tumor xenografts in response to fractionated radiation therapy, and to determine the feasibility of differentiating treatment response from failure. Additionally, we sought to determine whether classifier models based on Raman spectral markers could be used to distinguish between untreated radiation-resistant and sensitive tumors. To accomplish our goals, we used two sets of radiation-sensitive and radiation-resistant cell lines. First, we employed a recently developed matched model of radiation resistance [27], wherein a radiation-resistant clonal population of cells (rA549) was generated from parental A549 lung cancer cells. Second, we used HNSCC cell lines – UM-SCC-22B (UM22) and UM-SCC-47 (UM47) – for which radiation resistance and sensitivity have been established in previous studies [28]. Raman spectroscopic mapping of excised tumor xenografts (control and radiated tumors) grown from all four cell lines revealed consistent compositional alterations based on tumor type and in response to a radiation dose of only 2 Gy. Using multivariate curve resolution-alternating least squares (MCR-ALS), we translated the spectral information to uncover changes in lipid, collagen, and glycogen content. Data from both lung and head and neck (henceforth referred to as HN) tumors show consistently higher changes in lipid and collagen content in radiation-sensitive tumors that were treated with radiation compared with their radiation-resistant counterparts. Definition of the tumor phenotypes in terms of quantitative spectral features corresponding to key biomolecules also enabled the development of classifier models that exhibit high accuracy in discriminating between radiation-resistant and sensitive tumors.

Furthermore, our use of an isogenic radiation-resistant clone allowed, to the best of our knowledge, the first determination of discriminative Raman features in untreated tumors thereby offering fresh insights into specific molecular roles underlying intrinsic radiation resistance. Taken together, our findings highlight the potential of Raman spectroscopic imaging as a label-free, non-ionizing tool whose in vivo translation would permit monitoring of therapeutic effects with finer temporal resolution than is possible at the present time and potentially enable stratification of radiation-resistant patient cohorts.

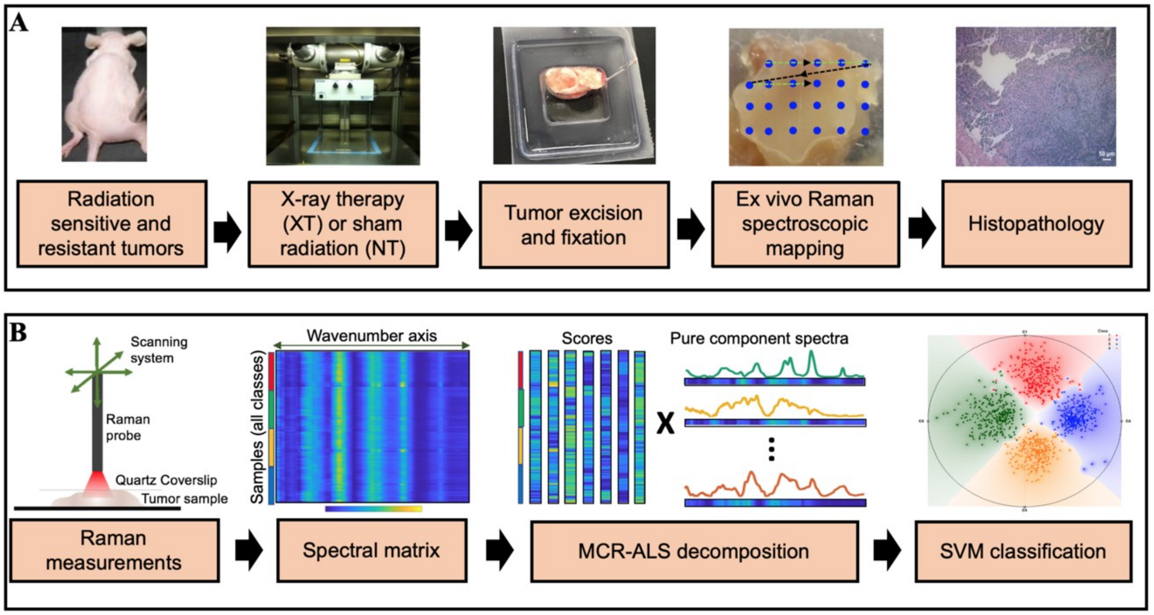


Figure 5.1: Raman spectroscopic study of radiation response and resistance. Overview of the (**A**) experimental and (**B**) data analysis workflow for Raman spectroscopic mapping in lung and head and neck tumor xenografts, of differential intrinsic radioresistance, subjected to radiation therapy. Details of the individual steps are provided in the methods section.

5.2 Materials and Methods

5.2.1 Cell culture

Human lung carcinoma A549 cells were purchased from American Type Culture Collection (ATCC; CCL185) and were authenticated using short tandem repeat (STR) profiling. A549 cells were grown in Ham's F-12K (Kaighn's) Medium mixed with 10% (v/v) fetal bovine serum and 1% (v/v) penicillin-streptomycin. These cells were irradiated at an average dose of 2.2 Gy every three days using an orthovoltage X-ray irradiator (CP-160, Faxitron X-Ray Corp. Wheeling, IL) for a cumulative dose of 55 Gy (25 fractions) to create the radiation-resistant cell clones (rA549) [27]. UM-SCC-22B and UM-SCC-47 were purchased from EMD Millipore and cultured in a mixture of Dulbecco's Modified Eagle Medium (DMEM), 10% Fetal Bovine Serum (FBS), 1% Penicillin-Streptomycin, 1% non-essential amino acids (NEAA), and 1% L-Glutamine. All cell lines used in this study tested negative for mycoplasma and were authenticated using short tandem repeat (STR) profiling.

Fractionated therapy of tumor xenografts. A schematic representation of this study design is presented in **Fig. 5.1A**. All animal studies were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Arkansas (Protocols 16022 and 18061). Athymic nu/nu mice were injected with a subcutaneous bolus of cells suspended in 100 µl of serum and media-free saline (10 million for A549 and rA549 cells, 2 million for UM-SCC-22B and UM-SCC-47) to grow tumor xenografts. Once tumor volume reached 200 mm^3, mice were randomized to either radiation (XT) or control (NT) groups, as presented in **Table 5.1**. Fractionated radiation therapy was administered using an X-RAD 320 biological irradiator (Precision X-Ray, North Branford, CT) as four 2 Gy fractions delivered over two consecutive weeks (total dose of 8 Gy), as described by others [28]. All animals completed the treatment. Tumor volumes were monitored using Vernier calipers, and tumors were excised when the majority of untreated control tumors had reached 1500 mm3 (approx. 35-50 days after treatment commenced). Tumor volume was calculated according to the equation V = (π/6) x (length) x (width) x (height). A comparison of tumor xenograft volumes is presented in Fig. SF1. After excision, tumors were embedded in OCT mounting medium, snap-frozen, and stored at -80 ^{o}C.

Table 5.1: Cell lines used to generate tumor xenografts in the study

Lung Tumors			**Head and neck tumors**		
Tumor group	*Class label*	*Number of tumors*	*Tumor group*	*Class label*	*Number of tumors*
A549-NT	A549-NT	5	UM-SCC-22B-NT	UM22-NT	6
A549-XT	A549-XT	4	UM-SCC-22B-XT	UM22-XT	7
rA549-NT	rA549-NT	5	UM-SCC-47-NT	UM47-NT	9
rA549-XT	rA549-XT	5	UM-SCC-47-XT	UM47-XT	10
Total number of tumors		**19**	**Total number of tumors**		**32**

5.2.2 Raman spectroscopy

The frozen tumors were thawed and fixed in 10% neutral buffered formalin prior to making Raman measurements. The fixed tumors were rinsed in PBS and sandwiched between a quartz cover slip and aluminum block to maintain a constant distance between the tissue and probe. The flattened tumors were scanned using a fiber-optic probe-based portable Raman spectroscopy system [16, 29]. Briefly, the custom-built system consists of an 830 nm diode laser (500 mW maximum power, Process Instruments) as the excitation source, and a combination of a spectrograph (Holospec f/1.8i, Kaiser Optical Systems) and a thermoelectrically (TE)-cooled CCD camera (PIXIS 400BR, Princeton Instruments) for recording spectra. Laser delivery and collection of back-scattered light was achieved via a lensed fiber-optic bundled Raman probe (probe diameter: 2 mm; Emvision LLC) with an estimated tissue sampling volume of 1 mm^3. The laser power at the tumor samples was maintained at ~ 20 mW in this study. Acquisition time for each spectrum was 10s (10 accumulations of 1 second each to prevent saturation of CCD). Tissue dehydration due to laser exposure was prevented by intermittent addition of PBS. Raster scanning of the probe using motorized translation stages (T-LS13M, Zaber Technologies Inc., travel range: 13 mm) and spectral acquisition were performed on each side of the flattened tumors (~ 100 mm^2) using a LabVIEW interface. About 4100 and 7000 spectra were acquired from the nineteen lung and thirty-two HN tumor xenografts, respectively.

5.2.3 Data analysis

Figure 5.1B illustrates the data analysis workflow in this study. All the data analysis in the present study was carried out using scripts written in MATLAB 2017a (Mathworks, Natick, MA, USA) environment unless otherwise stated. The wavenumber axis of the Raman system was calibrated using 4-acetamidophenol. The fingerprint wavenumber region (600 - 1800 cm-1) was chosen for further analysis. The Raman spectra recorded from the tumors were subjected to a fifth order best-fit polynomial-based fluorescence removal and cosmic ray removal using median filtering. The spectra were then vector normalized (such that their

Euclidean norm is set to unity) to minimize the effects of potential variations in laser power at the sample. The normalized spectra were used without any spatial averaging in the analysis. Multivariate curve resolution - alternating least squares (MCR-ALS) was employed to recover the pure spectral profiles of the chemical constituents of the tissue specimen without a priori information of the composition of the specimen [30]. The decomposition is achieved through an iterative optimization routine under non-negativity constraint on pure spectral (loadings) and concentration (scores) matrices. The non-negativity constraints enable us to interpret the unresolved specimen spectra in the form of loadings that represent spectra of pure (or enriched) biochemical components and the corresponding scores that provide a measure of abundance of the particular component. Additionally, spectral equal length constraint is imposed on the pure spectra to facilitate comparison of corresponding scores across the classes (treatment groups). The normalized scores corresponding to each key biological constituent were compared across different classes through box and whisker plots with conventional thresholds. The spatial heterogeneity in the score maps of major biological constituents for each tumor was quantified in terms of distributional homogeneity index (DHI), as defined elsewhere in the literature [31]. The significance of differences in medians of constituent scores across studied classes were assessed based on two-sided Wilcoxon rank sum test statistics. A conventional criterion of p-value less than 0.05 was used to consider the medians different. The differences between the groups were quantified in terms of effect size calculated using the Wendt formula for rank biserial correlation [32].

Support vector machine (SVM) was used to develop a decision algorithm to identify radiation treatment response and to predict resistant phenotype prior to treatment. SVM is a nonlinear classification method wherein classification is achieved by solving a constrained quadratic optimization problem to obtain separating boundaries between the classes in higher dimensional spaces [33]. In this study, the LIBSVM library [34] was used to develop a C-SVM classifier. The background-corrected spectra were used along with tumor group labels for each group studied, without any spatial averaging. A radial basis function kernel with a Gaussian envelope was employed to enable nonlinear mapping of the input feature space, and the optimal C-SVM parameters (i.e. cost and kernel parameter gamma) were selected using a k-fold cross validation-based grid search algorithm. A leave-one-mouse-out analysis was conducted in which spatially distinct spectra belonging to each mouse were eliminated from the training dataset and the resulting binary SVM classifiers for pairs of classes of interest were tested using the spectra belonging to the left-out mouse. Each mouse specimen was assigned an overall predicted class label if more than 90% of its spectra were predicted as belonging to that class; otherwise, the specimen was labeled as unclassified if the desired confidence level was not achieved. Randomized class equalization was performed iteratively prior to implementing SVM classification in order to avoid skewing the model due to varying class sizes.

5.2.4 Histopathology

The tumors were stored in 70% ethanol after acquisition of Raman spectra and submitted to the Phenotyping and Pathology Core at Johns Hopkins Medical

Institutions. The formalin-fixed tumors were embedded in paraffin and sectioned serially onto glass slides for histology. Hematoxylin and eosin (H&E) staining, Masson's trichrome staining for collagen and Periodic acid Schiff (PAS) staining (without Hematoxylin counterstain) for glycogen were performed by the Core according to standard protocols. The stained slides were imaged using a Leica DMi8 inverted optical microscope. In addition, Oil Red O staining for lipids was performed on frozen tumor sections according to standard IHC protocols and imaged using a Nikon fluorescence microscope.

5.3 Results

To capture the tumor heterogeneity and variance arising from differential response to treatment, lung and HN tumors in each group – radiation treated (XT) and controls (NT), were mapped to obtain spatially distinct Raman spectra from each specimen (**Fig. 5.2A**). Each map had an average of 218 spectra (ranging between 50 and 334 spectra depending on the size of the tumor). Except for this visualization (**Fig. 5.2A**), the spectra collected from each tumor specimen were treated individually without any spatial averaging in all analyses. The spectra across all the classes show prominent peaks at 1045 cm^{-1} (glycogen), 1256 cm^{-1} (glycogen), 1301 cm^{-1} (CH vibration in lipids), 1448 cm^{-1} (CH_2 bending modes in lipids and collagen), and 1656 cm^{-1} (amide-I feature of proteins with potential contribution from C=C stretching in lipids). To discern possible molecular differences, the spectral datasets were decomposed into key compositional biomolecular signatures that were compared across the different groups.

To achieve this decomposition, we performed MCR-ALS with 7 components to obtain a loadings matrix containing the ‘pure component’ basis spectra and a scores matrix containing the weights of each of the seven components for all the spectra in the dataset. **Figures 5.2B** and **5.2C** illustrate the relevant MCR loadings that present Raman features corresponding to key tissue constituents in the lung and HN tumor cohorts, respectively. The remaining loadings, that do not show direct correspondence to the vibrational signature of a prominent molecular constituent or stem from the presence of formalin (spectral contaminant in the tissue specimen), are provided in the **Fig. S5.2** (Appendix). As seen in **Fig. 5.2B**, spectrum B1 shows prominent peaks at 1078 cm^{-1}, 1266 cm^{-1}, 1301 cm^{-1}, 1442 cm^{-1} and 1654 cm^{-1} that are characteristic of lipids, specifically triglycerides. Spectral pattern B2 resembles the Raman spectral profile of glycogen with peaks at 708 cm^{-1}, 940 cm^{-1}, 1044 cm^{-1}, 1078 cm^{-1} and 1256 cm^{-1}. Furthermore, B3 has spectral features corresponding to nucleic acids at 790 cm^{-1}, 812 cm^{-1} and 1082 cm^{-1}, while loading B4 has peaks at 851 cm^{-1}, 928 cm^{-1}, 1040 cm^{-1}, 1251 cm^{-1}, 1315 cm^{-1}, 1453 cm^{-1} and 1661 cm^{-1}, which are characteristic of collagen. The loadings derived from the HN tumor dataset C1, C2 and C3 presented features similar to the loadings B1, B3 and B4, respectively. The detailed peak allocations of all the features of the 7 loadings derived from lung and HN tumor datasets have been tabulated in **Tables ST5.1** and **ST5.2** (Appendix), respectively. It is worth noting that the MCR decomposition of HN tumors did not reveal a glycogen-rich loading. The minor spectral features that stem from the use of formalin as a fixative

– at 1490 cm^{-1} and 1040 cm^{-1} – present themselves in a single MCR loading, thus indicating that the effects of formalin fixation may be digitally removed [35].

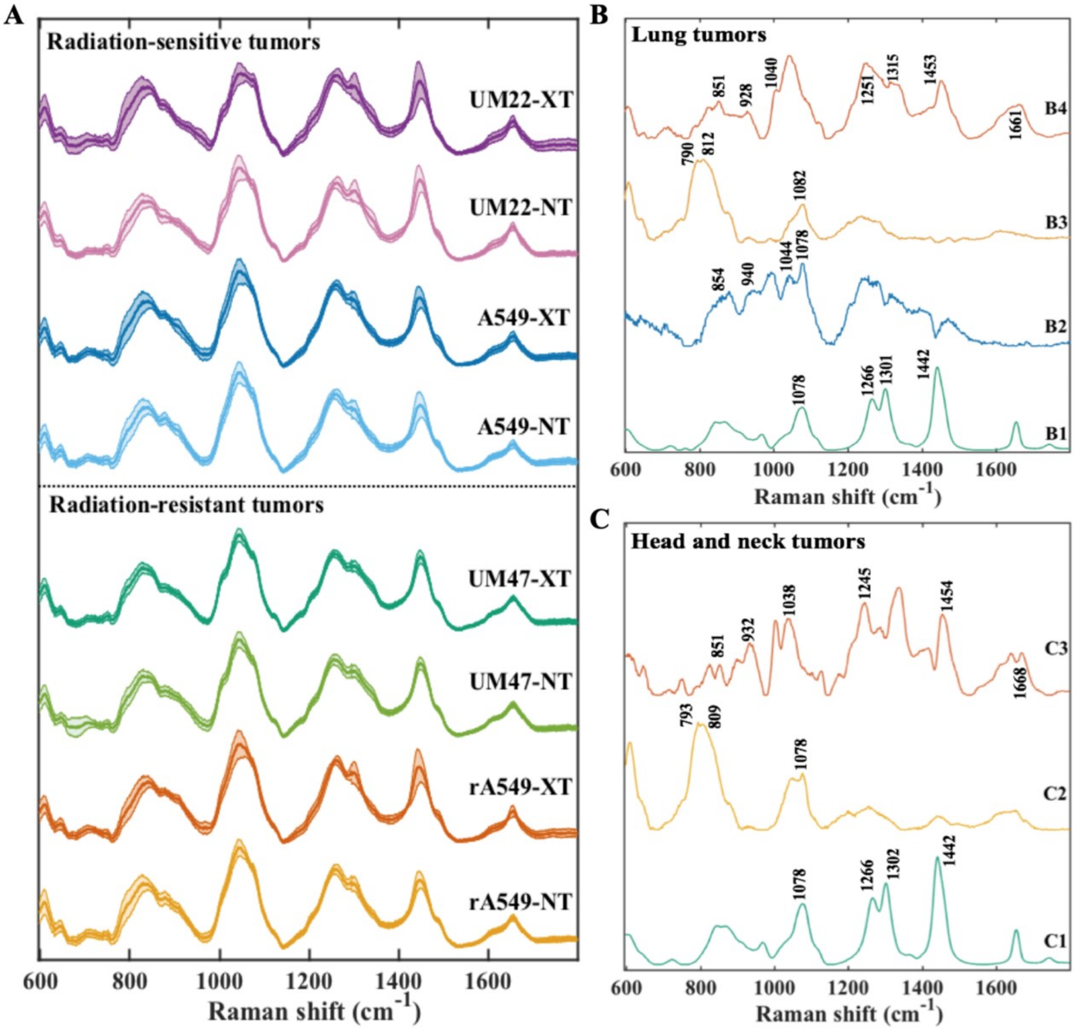


Figure 5.2: Raman spectra of radiation-resistant and sensitive tumors. (**A**) Mean Raman spectra (with the shadow representing 1 standard deviation) acquired from lung tumors derived from parental A549 and radiation-resistant (rA549) cells and head and neck tumors derived from radiation-sensitive UM-SCC-22B and radiation-resistant UM-SCC-47 cells. Each of the types were either exposed to fractionated radiation (XT) or sham radiation (NT). (**B**) A subset of relevant MCR loading vectors derived from the spectra of lung tumors belonging to all the four study classes. The spectra B1 through B4 represent lipid-rich, glycogen-rich, nucleic acid-rich and collagen-rich loadings, respectively. (**C**) A subset of relevant MCR loading vectors derived from the spectra of tumors belonging to head and neck tumor dataset. The spectra C1 through C3 represent lipid-rich, nucleic acid-rich and collagen-rich loadings, respectively.

Representative MCR score maps (abundance maps) of the observed relevant loadings across the treatment groups have been provided in **Fig. S5.3** (Appendix). To characterize the spatial heterogeneity in biochemical composition of the studied tumors, we measured the distributional homogeneity index (DHI) for each MCR score map. DHI is a measure of deviation of the spatial distribution in compositional maps from their randomized counterparts [31]. Thus, DHI is positively correlated with heterogeneity in abundance maps, with a value of one representing homogeneous (or randomized) distribution and higher values indicative of localization of constituents. Therefore, from the observed large values of DHI (**Fig. S5.3**), it is evident that the abundance maps for relevant loadings obtained for lung and HN tumors exhibit substantial spatial heterogeneity in the tumor samples.

The primary objectives of our study were to investigate differences in radiation-induced microenvironmental changes between resistant and sensitive tumors, and whether such molecular alterations were consistent in tumors derived from two different sites – lung and HN. Accordingly, NT vs. XT comparisons for each of the 4 cell lines were performed to evaluate the differential response of sensitive (A549, UM-SCC-22B) and resistant (rA549, UM-SCC-47) tumors. A secondary objective was to determine if these molecular features could distinguish between untreated resistant and sensitive tumors. Since the HN tumors stemmed from different cell lines, this specific comparison was only performed in the lung cohort (A549-NT vs rA549-NT) to identify intrinsic differences underlying radiation resistance. Here, we used density scatter plots as they allow better visualization of large datasets by avoiding overlap in regions of high density. **Figures 5.3A-C** present three-dimensional (3D) density plots using the normalized MCR-ALS scores corresponding to the lipid-rich, glycogen-rich and collagen-rich loadings of A549 and rA549 tumors. Similarly, two-dimensional (2D) density plots with only lipid-rich and collagen-rich loadings are shown in **Figs. 5.3D-E** for the HN tumors. Both sets of density plots underscore expected tumor heterogeneity and critically, the presence of group-specific local spectral clustering, which is reflected in the higher density of co-located points (circled in the figure).

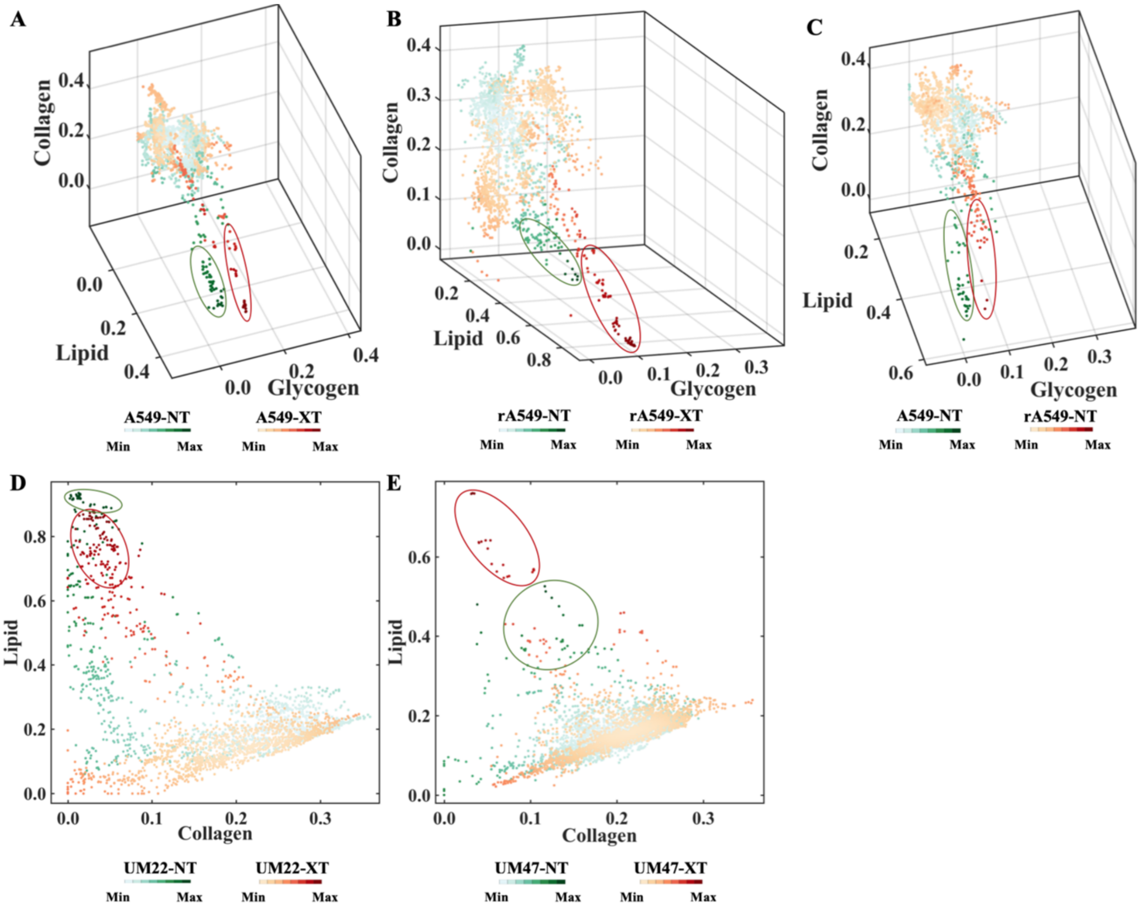


Figure 5.3: Qualitative visualization of MCR-ALS scores of Raman spectra. (**A-C**) Three-dimensional density plots showing the distribution of normalized scores of lipid-rich, collagen-rich and glycogen-rich MCR-ALS loadings showing radiation induced differences in sensitive lung tumors (A549-NT vs A549-XT), radiation induced differences in resistant lung tumors (rA549-NT vs rA549-XT) and pre-radiation differences between sensitive and resistant lung tumors (A549-NT vs rA549-NT), respectively. (**D-E**) Two-dimensional density plots showing the distribution of normalized scores of lipid-rich and collagen-rich MCR-ALS loadings showing radiation induced differences in sensitive head and neck tumors (UM22-NT vs UM22-XT) and radiation induced differences in resistant head and neck tumors (UM47-NT vs UM47-XT), respectively. The class specific clustering in high density regions are circled.

To quantitatively study the differences across the treatment groups, we compared the normalized MCR-ALS scores of lipid-rich, collagen-rich and glycogen-rich loadings. We observed an increase in lipid, collagen, and glycogen levels for both sensitive (**Fig. 5.4A**) and resistant (**Fig. 5.4B**) lung tumors that were treated with radiation, with a much larger increase in the lipid-rich and collagen-rich signatures in the radiation-sensitive tumors. To examine intrinsic radioresistance, we also compared the scores of these biomolecular components between untreated A549 and rA549 tumors (**Fig. 5.4C**). The MCR-ALS scores point to higher lipid and collagen content but lower glycogen content in the resistant tumors compared with the sensitive tumors. Furthermore, we observed very similar effects in the HN tumors with a greater increase in lipid and collagen levels in the radiation-sensitive UM-SCC-22B (**Fig. 5.4D**) compared to the radiation-resistant UM-SCC-47 tumors (**Fig. 5.4E**). Glycogen, as mentioned previously, was not observed as a significant feature in the HN tumors. Notably, the greater changes in sensitive tumors post-radiation are also borne out by the effect sizes for lipid and collagen content, which are consistently higher for the sensitive tumor cohorts (≥ 0.35 in A549 and UM-SCC-22B; ≤ 0.16 in rA549 and UM-SCC-47).

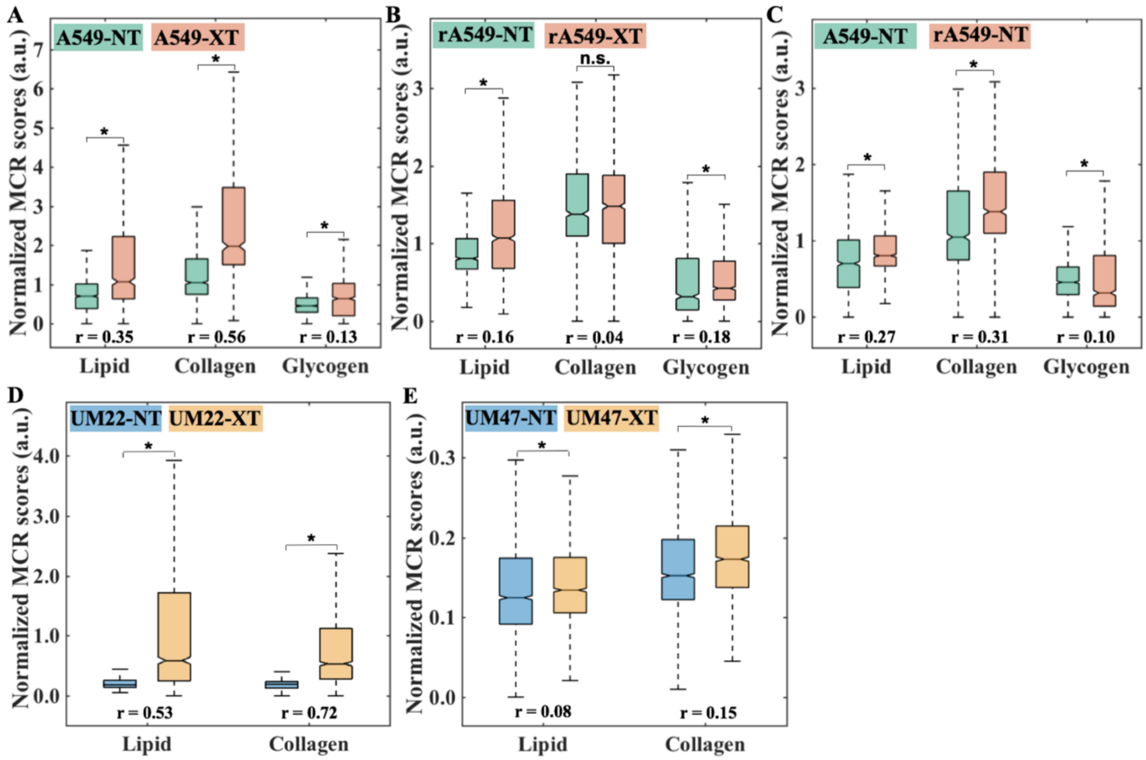


Figure 5.4: Quantitative MCR-ALS analysis of Raman spectra. (**A-C**) Boxplots of normalized scores of lipid-rich, collagen-rich and glycogen-rich MCR-ALS loadings showing radiation induced differences in sensitive lung tumors (A549-NT vs A549-XT), radiation induced differences in resistant lung tumors (rA549-NT vs rA549-XT) and pre-radiation differences between sensitive and resistant lung tumors (A549-NT vs rA549-NT), respectively. The differences in the scores of lipid and glycogen loadings are statistically significant (indicated by * and n.s. otherwise) at $p < 0.001$ level (Wilcoxon rank sum test) for each of the three comparisons (**A-C**), whereas the differences in the scores of collagen loadings are statistically significant only for the comparisons in (**A**) and (**C**). (**D-E**) Boxplots of normalized scores of lipid-rich and collagen-rich MCR-ALS loadings showing radiation induced differences in sensitive head and neck tumors (UM22-NT vs UM22-XT) and radiation induced differences in resistant head and neck tumors (UM47-NT vs UM47-XT), respectively. The differences in the scores of lipid and collagen loadings are statistically significant (indicated by *) at $p < 0.001$ level (Wilcoxon rank sum test) for both the comparisons. The effect size (r), characterizing magnitude of differences between groups, is provided for each comparison.

While the comparison of MCR scores provides a starting point for delineating the molecular mediators of treatment response/resistance and assessing the predictive power of the spectroscopic data, comparison of the individual component scores alone may not provide a robust diagnostic framework, especially to classify prospective samples. Therefore, we developed decision models based on support vector machines (SVM), a supervised classification method that can deal with ill-posed problems and lead to unique global models [33]. We conducted a leave-one-mouse-out analysis which involved training three separate binary SVM classifiers for each tumor type (lung and HN tumors), corresponding to the three sets of comparisons – RS-NT vs RS-XT, RR-NT vs RR-XT and RS-NT vs RR-NT, respectively. RS and RR indicate radiation-sensitive and radiation-resistant tumor xenografts, respectively. **Table 5.2** shows the aggregated number of mice accurately classified, unclassified, and misclassified (as determined by the criteria detailed in the Methods section) for each of the three comparisons. (The tumor model-specific decomposition is provided in Appendix **Tables ST5.3** and **ST5.4**.)

The leave-one-mouse-out protocol provides satisfactory predictions in all cases with an overall misclassification rate of only *ca.* 3%. While slightly higher unclassification rates were noted for the lung tumor dataset, incorporation of a larger cohort of animals in the HN tumor data allowed significant reduction in the same. In the latter set of HN tumors, slightly increased unclassification rate (albeit with zero misclassification) was observed for the comparison between treated (UM47-XT) and untreated (UM47-NT) radiation-resistant HN tumors. We attribute this increased unclassification to smaller effect size observed in MCR-ALS-based univariate comparisons and the high classification threshold used in our leave-one-mouse-out analysis. For example, relaxing the threshold to 80% level reduced the unclassification rates of comparison between UM47-NT and UM47-XT significantly (2/19 mice compared to 7/19 mice at 90% level).

Table 5.2: Results of binary leave-one-mouse-out SVM analyses

		Number of mice classified accurately, unclassified and misclassified			
		RS-NT	**RS-XT**	**RR-NT**	**RR-XT**
Binary SVM comparisons	**RS-NT vs RS-XT**	(11+0+0)/11	(8+2+1)/11	-	-
	RR-NT vs RR-XT	-	-	(9+5+0)/14	(12+2+1)/15
	RS-NT vs RR-NT	(10+1+0)/11	-	(12+2+0)/14	-
	RS-XT vs RR-XT	-	(10+0+1)/11	-	(14+1+0)/15

Finally, to verify the lack of spurious correlations in the dataset [36], we repeated our leave-one-mouse-out analysis using the same spectral dataset, but with randomly assigned class labels instead of their original labels for each comparison. Average correct classification rate of ca. 57% for both lung tumor and HN tumor datasets, were obtained (comparable to the random likelihood of selection of the true class label – 50%). Taken together, the results of the SVM-derived classifier model studies demonstrate the utility of the Raman spectroscopic data in capturing distinct radiobiological responses in radiosensitive and radioresistant lung and HN tumor xenografts.

5.4 Discussion

A fundamental principle of personalized medicine is to design treatment strategies that tackle the biological heterogeneity characteristic of cancer in order to achieve maximal tumor control while minimizing toxicity. The lack of suitable imaging tools, which can identify patients unlikely to benefit from radiation and perform frequent response monitoring to better inform treatment doses and fractionation schemes, remains a major impediment in customizing radiotherapy. In contrast to existing clinical technologies, optical spectroscopy offers a non-invasive or minimally invasive route to providing real-time evaluation of treatment response based on functional and biomolecular changes in the tumor microenvironment. In this study, we demonstrate the utility of using label-free Raman spectroscopy in conjunction with chemometric analysis to reveal distinct biomolecular changes in radiation-resistant and sensitive tumors when subjected to fractionated, clinically-relevant radiation doses. Specifically, MCR-ALS analysis reveals consistent differences in lipid and collagen content post-radiation in the microenvironment of lung and head and neck tumor xenografts with especially pronounced changes in the radiation-sensitive cases. By leveraging SVM-derived classifiers, we are also able to differentiate between vibrational signatures recorded from untreated radiation-sensitive and radiation-resistant tumors indicating the potential for future Raman spectroscopic application to not only monitor but also predict radiation response in individuals.

The specific biomolecular features derived using MCR-ALS analysis have been previously studied in the context of cellular response to radiation. Hypoxia and its transcription factor, hypoxia-inducible factor (HIF-1), have been shown to promote extracellular matrix (ECM) remodeling and play an important role in promoting fibrosis [37, 38] and collagen biogenesis [39, 40]. Indeed, our recent investigation of cellular metabolism showed a significant increase in HIF-1α expression after radiation in both the A549 and rA549 cells [41]. We reason that the radiation-induced increase in HIF-1 content is, in part, responsible for promoting collagen deposition in the A549 and rA549 tumors. The increased collagen content may also be explained by the actions of growth factors, such as TGF-β, which are recruited in response to HIF-1α-stimulated macrophage accumulation [42]. Overexpression of TGF-β serves as a chemoattractant for the recruitment of fibroblasts, and may drive the increase in collagen-rich MCR scores upon irradiation for both sensitive and resistant tumors [43].

Previous work has established that de novo lipogenesis protects cancer cells from external insults, such as oxidative stress, and that inhibition of lipogenesis increases oxidative stress-induced cell death [44]. The increased lipid content observed in both groups of radiated tumors in our study could be attributed to such a cellular defense mechanism in response to radiation-induced oxidative stress. Although the exact mechanism for increased lipid content needs further investigation, studies have found elevated levels of fatty acid synthase (FASN) in radiation-resistant HN cancer cells [45]. Furthermore, inhibition of FASN decreased cellular survival of these radiation-resistant cancer cells. FASN is a key player in lipogenesis and has also been shown to be a prognostic indicator of radiation resistance in clinical nasopharyngeal carcinoma [46]. Taken together, these studies highlight the potential of lipids to serve as a powerful biomarker of radiation resistance.

Our findings of increased glycogen content in response to radiation in the radiation-resistant rA549 tumors are consistent with a recent in vitro study by Matthews et al. that reported an increase in radiation-induced glycogen in the relatively radiation-resistant MCF7 breast cancer and H460 lung cancer cell lines [23]. Jirasek and colleagues have reported substantially increased glycogen content in radiated non-small cell lung tumor xenografts compared with non-radiated xenografts [26]. However, while our study also found significantly higher glycogen content in the radiation-sensitive A549 tumors, Mathews and colleagues found no changes in glycogen content in the radiation sensitive LNCaP prostate cancer cells. These differences could be principally attributed to the different nature of the measurement specimen (cells vs. tissue). The increase in intracellular glycogen in their study was attributed to the phosphorylation of glycogen synthase kinase (GSK-3β), a negative regulator of glycogen synthase, which mediates the final step of glycogen synthesis. Deactivation of GSK-3β activity through phosphorylation has been shown to play a critical role in the acquisition of radiation resistance in cancer cells [47]. Although a direct role for glycogen in conferring protection from radiation has not been established, the availability of increased glycogen reserves could provide cancer cells with glucose through glycogenolysis during radiation-induced oxidative stress. Glucose utilization through the pentose phosphate pathway can lead to the generation of glutathione, which is an important scavenger of radiation-induced free radicals. Our findings expand on these determinations by providing the first direct comparisons of glycogen levels in matched models of resistant and sensitive tumors. However, the smaller effect sizes observed in the lung tumors coupled with its absence in the HN tumors suggest that further investigations are necessary to evaluate the clinical utility of glycogen as a marker of treatment response.

To examine the histological basis of the Raman spectroscopic determinations, tumor sections were stained with hematoxylin and eosin (H&E), Masson's trichrome, periodic acid-Schiff (PAS) and oil red O (ORO) – for cellular morphology, collagen, glycogen and lipid, respectively (**Fig. 5.5A-L** and **S5.4**). While tumor morphology was found to be largely similar across all tumor groups, H&E-stained images identified high levels of necrosis as well as fibrosis in the tumor groups subjected to radiation therapy (XTs) compared with the untreated control tumors (NTs). Specifically, necrosis and fibrosis were found in all groups, and were correlated, with an increase in necrosis associated with an increase in fibrosis. Specifically, within the lung tumor group, the A549-NT and rA549-NT tumors demonstrated higher tumor burdens and lower levels of necrosis and fibrosis. Within the HN tumor group, the lowest levels of necrosis were observed in the 22B-NT tumors while the highest levels were observed in the 47-XT group. Both 22B-XT and 47-NT had similar and intermediate levels of necrosis and fibrosis. This is likely due to the bulkier tumors observed in the 47-NT group. This tumor overgrowth is often associated with a degenerative type of necrosis secondary to ischemia. These histopathological results were largely consistent with the findings from Raman spectral analysis. Furthermore, using bright field images of the Masson's trichrome and PAS stained slides, we observed increased levels of collagen and glycogen after radiation in both the sensitive and resistant tumors. In addition, collagen content was noticeably higher in the resistant tumors prior to radiation compared with the sensitive tumors. The histological images for the HN tumors have been provided in **Fig. S5.4** (Appendix).

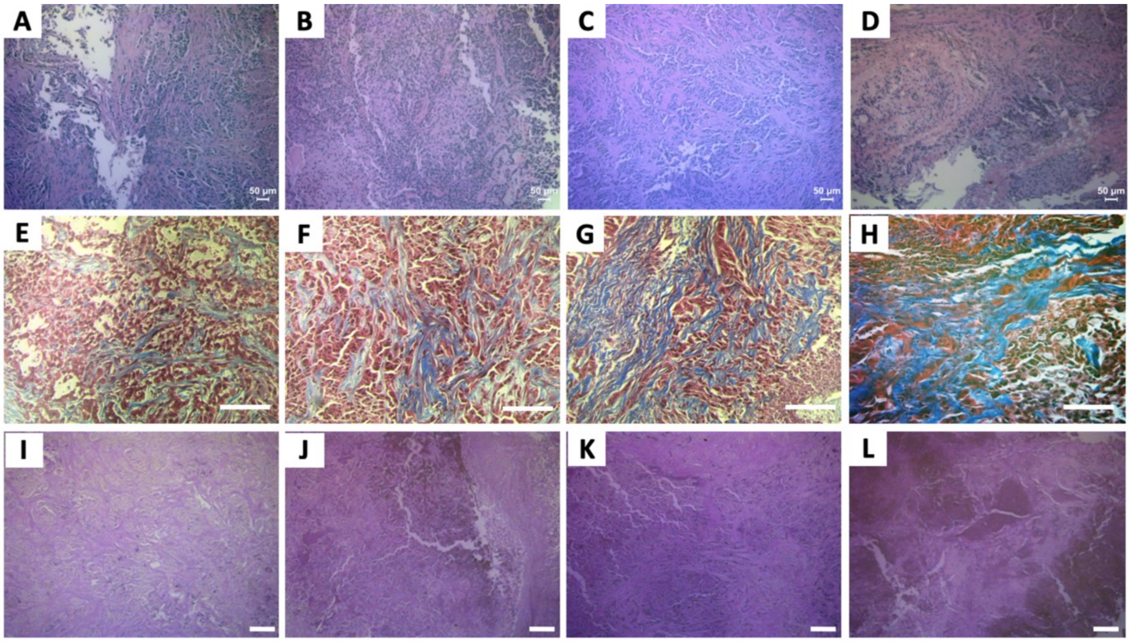


Figure 5.5: Histologic assessment of radiation sensitivity and resistance. Top (**A–D**), middle (**E–H**) bottom (**I-L**) panels display representative microscopic images of H&E, Masson's trichrome and PAS stained slides, respectively. The columns of panels from left to right - (**A, E and I**), (**B, F and J**), (**C, G and K**) and (**D, H and L**), respectively, represent fields of view from tumors belonging to the treatment groups - A549-NT, A549-XT, rA549-NT and rA549-XT. The scale bars in panels **A-D** represent 50 µm and panels **E-L** represent 100 µm.

In summary, we have used Raman spectroscopic mapping for quantitative assessment of the molecular composition of lung and HN tumors subjected to radiation therapy, and shown that such measurements offer a reliable, non-perturbative method to probe radiation-induced alterations. These findings represent, to the best of our knowledge, the first report comparing the microenvironmental response to radiation in tumor xenografts from different organ sites using optical spectroscopy. Together, our results provide promising evidence for the clinical translation of Raman spectroscopy to discern molecular markers of radiation response either prior to or during the early stages of treatment using fiber optic probes in accessible tumors.

Towards that goal, there are two major focal points of our future investigations. First, the clinical radiation dose of 2 Gy that was used in this study will be delivered on successive days as is usually performed in the clinic. The treatment regimen used here is similar to previous approaches used to establish radiation sensitivity and resistance in tumor xenograft models [28]. Second, radiation-induced microenvironmental changes were evaluated ex vivo from excised tumors. Our evaluation of differences related to intrinsic radiation resistance were conducted on the untreated control tumors. While we observed appreciable differences between the A549-NT and rA549-NT tumors that can be attributed to radiation resistance, these results do not exactly predict if these tumors would go on to respond or fail treatment. The results presented here provide an opportunity to further explore the prediction of long-term treatment response based on measurements made prior to commencing treatment in radiation-naïve tumors in vivo. Our next study would involve in vivo pre-treatment measurements on tumors as well continuous measurements during treatment administered on successive days to enable longitudinal treatment monitoring. Our recent work using diffuse reflectance spectroscopy identified changes in tumor oxygenation in the A549 and rA549 tumors within 48 hours post-radiation; however, minimal or no differences in oxygenation were observed at the time of tumor excision [48]. Therefore, it is possible that the magnitude of radiation-induced biomolecular changes, as sensed by in vivo Raman measurements, will be greater immediately after radiation therapy. Moreover, the fabrication of appropriate probes [49] as well as the emergence of vibrational spectroscopic imaging systems that are already being adopted in clinical studies [50] indicate that translation of our proposed approach is feasible. Hence, based on our current findings as well as these technological developments, we envision that Raman measurements will be employed in the near future to guide treatment planning based on the inclusion of vibrational spectral profiles of a patient’s tumor.

5.5 Appendix

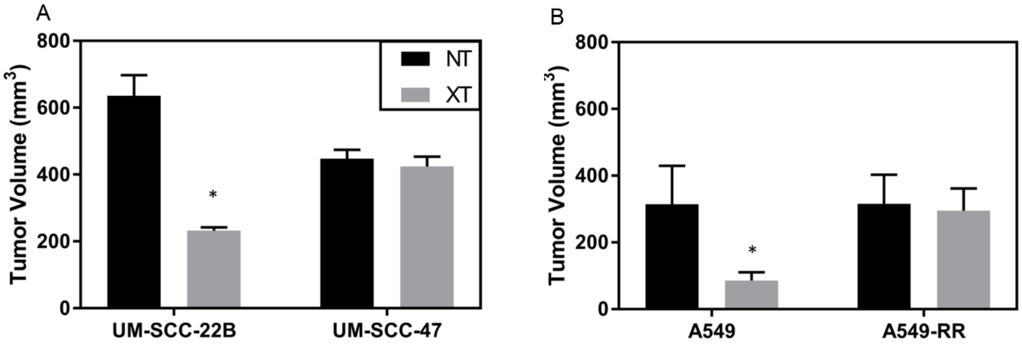


Figure S5.1: Tumor growth assays in response to radiation therapy. For (**A**) HNSCC tumors and (**B**) Lung tumors. Tumors in the XT group were radiated using the radiation schedule described in the methods section (4 x 2 Gy) at tumor volumes between 100 – 200 mm^3. NT groups received sham radiation. There were significant differences in tumor volume between NT and XT groups in the UM-SCC-22B and A549 tumors. Differences in tumor volume were evaluated 35-50 days post-treatment. * indicates significant differences at $p < 0.05$.

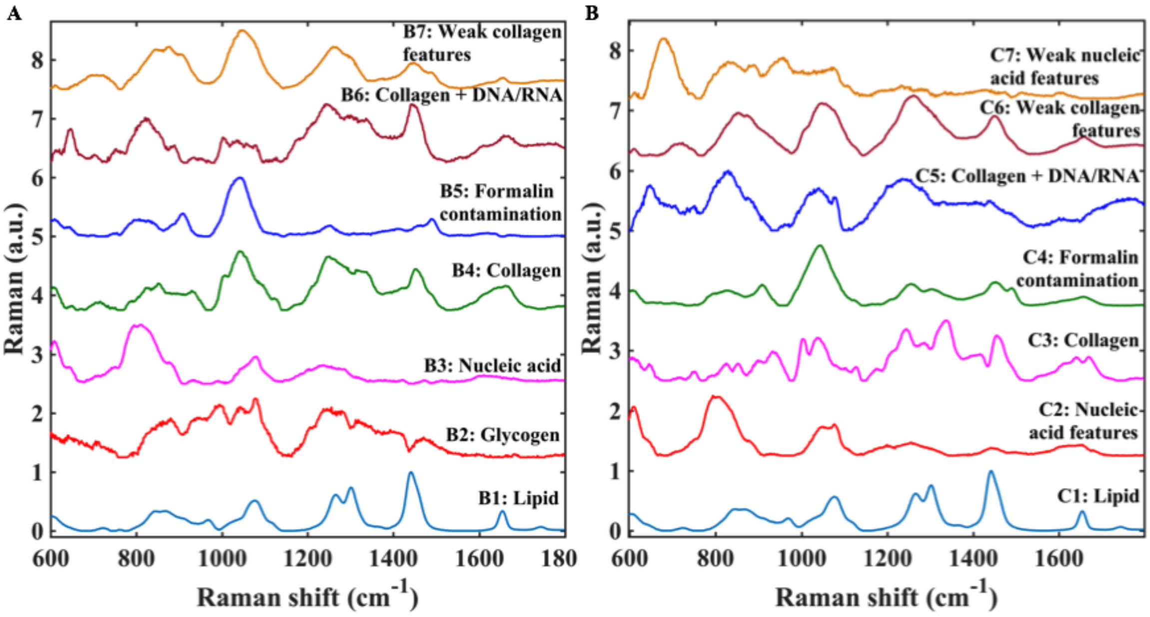


Figure S5.2: Complete set of MCR-derived pure component spectra. The complete set of MCR-derived pure component spectra are provided here for – (**A**) lung tumor, and (**B**) head and neck tumor datasets. The biological components of interest, whose pure spectra resembles those obtained through MCR-ALS analysis, are identified and marked alongside the spectra.

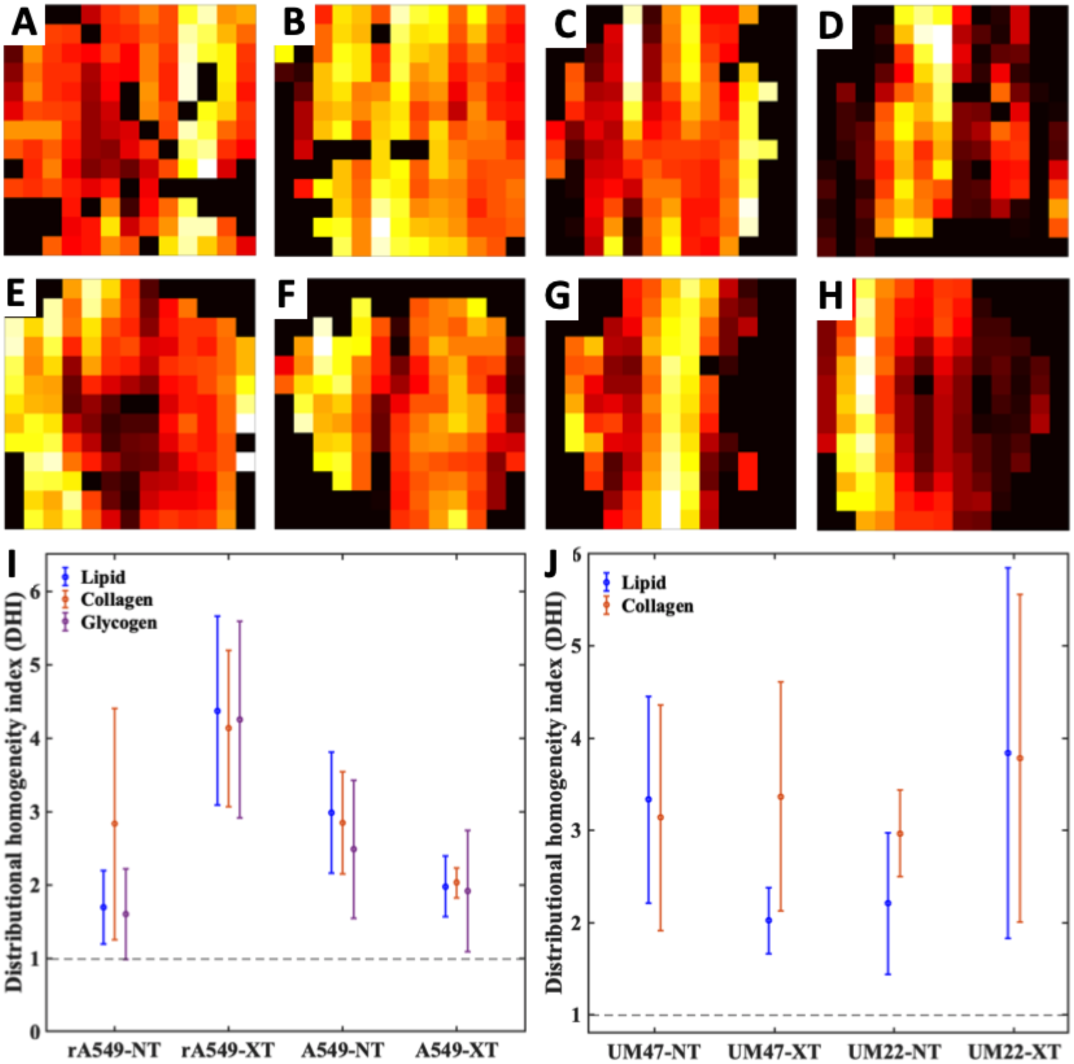


Figure S5.3: Heterogeneity in biochemical composition of tumors. Top (**A–D**) and middle (**E–H**) panels display representative MCR-ALS score maps of glycogen-rich and lipid-rich loadings respectively in lung and HN tumors. The columns of panels from left to right - (**A and E**), (**B and F**), (**C and G**) and (**D and H**), respectively, represent fields of view from tumors belonging to the treatment groups - RR-NT, RR-XT, RS-NT and RS-XT, where RR and RS are the radiation resistant and radiation sensitive groups respectively. Each pixel in the images in panels **A-F** spans an area of 1mm x 1mm. The bottom (**I and J**) panels show distributional homogeneity index plots calculated using relevant MCR-ALS loading score maps for tumors in lung and HN tumor datasets, respectively.

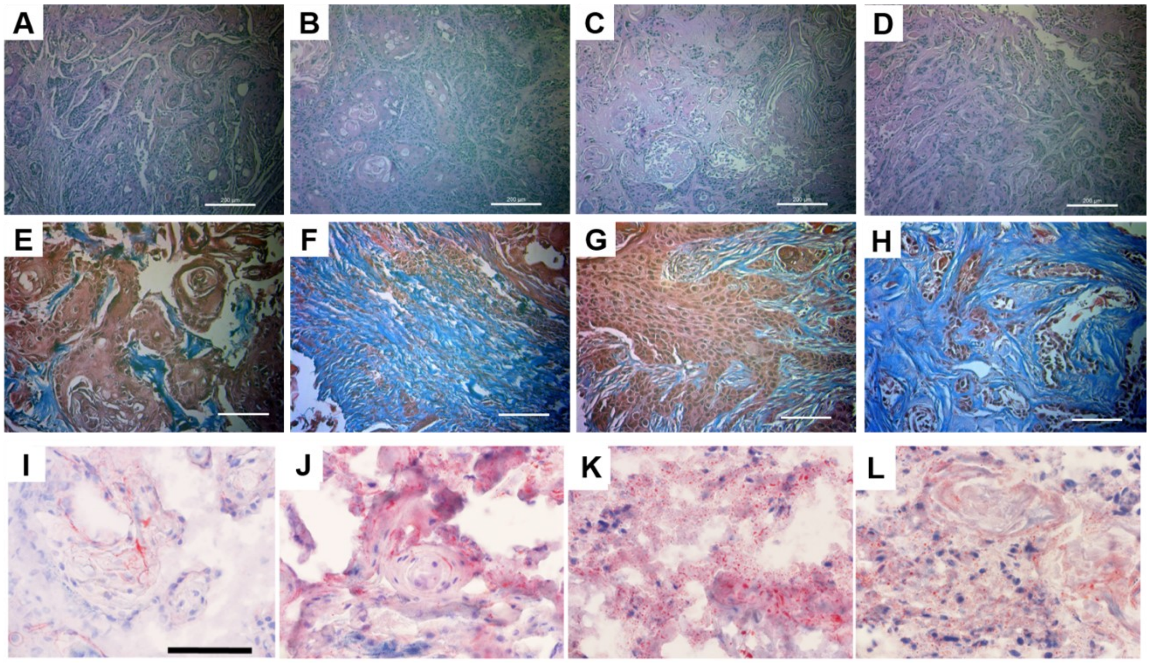


Figure S5.4: Histologic assessment for head and neck tumor dataset. Top (**A–D**), middle (**E–H**) and bottom (**I–L**) panels display representative microscopic images of H&E, Masson's trichrome, and Oil Red O stained slides, respectively. The columns of panels from left to right - (**A and I**), (**B and F**), (**C and G**) and (**D and H**), respectively, represent fields of view from tumors belonging to the treatment groups – UM22-NT, UM22-XT, UM47-NT and UM47-XT. The scale bars in panels **A-D** represent 200 µm and panels **E-L** represent 100 µm.

Table ST5.1: Table listing the peak assignments for all the MCR-derived component spectra derived from lung tumor dataset

Observed Raman peaks in the MCR loadings (cm^{-1})							Raman band assignment from literature
B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	
	708						Glycogen
		790					O-P-O stretching in DNA
		812			816		O-P-O stretching in DNA and RNA
			851				C-C stretch of proline in collagen
						878	C-C stretch of hydroxyproline in collagen
				908			Formalin contamination during tissue fixation
			928				C-C vibration in collagen backbone
	940						Glycogen
			1040				Proline in collagen
				1042			Formalin contamination during tissue fixation
	1044						Glycogen
						1047	Proline in collagen
	1078						Glycogen
1078							C-C stretch
		1082			1080		PO_2^- symmetric stretching in DNA
		1237					PO_2^- asymmetric stretching in DNA
					1242		Amide III in collagen
			1251				Amide III in collagen
				1251			Formalin contamination during tissue fixation
	1256						Glycogen
						1259	Amide III in collagen

1266							CH_2 in-plane deformation (Triglyceride)
1301							CH vibration (Triglyceride)
	1320						Glycogen
			1315				CH_3CH_2 twisting modes of collagen
					1337		CH_3CH_2 wagging modes of collagen and nucleic acids
1442							CH_2 bending mode (Triglyceride)
			1453			1448	CH_2 bending mode in collagen
				1489			Formalin contamination during tissue fixation
1654							C=C lipid stretch
			1661		1657	1656	α-helical structure of amide I in collagen

Table ST5.2: Table listing the peak assignments for all the MCR-derived component spectra derived from head and neck tumor dataset

Observed Raman peaks in the MCR loadings (cm^{-1})							Raman band assignment from literature
C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	
	790						O-P-O stretching in DNA
	806						O-P-O stretching in DNA and RNA
				820			O-P-O stretching in DNA and RNA
		852					C-C stretch of proline in collagen
					873		C-C stretch of hydroxyproline in collagen
			906				Formalin contamination during tissue fixation
		932					C-C vibration in collagen backbone
		1038					Proline in collagen
			1042				Formalin contamination during tissue fixation
					1047		Proline in collagen
1078							C-C stretch
	1076			1076		1074	PO_2^- symmetric stretching in DNA
				1235			Amide III in collagen
		1245					Amide III in collagen
			1256				Formalin contamination during tissue fixation
				1259	1262		Amide III in collagen
1266							CH_2 in-plane deformation (Triglyceride)
1302							CH vibration (Triglyceride)
			1304				Formalin contamination during tissue fixation
							CH_3CH_2 twisting modes of collagen

		1338				1335	CH_3CH_2 wagging modes of collagen and nucleic acids
1442							CH_2 bending mode (Triglyceride)
		1454			1448		CH_2 bending mode in collagen
			1451				Formalin contamination during tissue fixation
			1491				Formalin contamination during tissue fixation
1654							C=C lipid stretch
		1668			1657		α-helical structure of amide I in collagen

Table ST5.3: Results of binary leave-one-mouse-out SVM analysis for lung tumor dataset

		Number of mice classified accurately, unclassified and misclassified			
		A549-NT	**A549-XT**	**rA549-NT**	**rA549-XT**
Binary SVM comparisons	**A549-NT vs A549-XT**	(5+0+0)/5	(1+2+1)/4	-	-
	rA549-NT vs rA549-XT	-	-	(5+0+0)/5	(4+0+1)/5
	A549-NT vs rA549-NT	(5+0+0)/5	-	(3+2+0)/5	-
	A549-XT vs rA549-XT		(3+0+1)/4		(4+1+0)/5

Table ST5.4: Results of binary leave-one-mouse-out SVM analysis for head and neck tumor dataset

		Number of mice classified accurately, unclassified and misclassified			
		UM22-NT	**UM22-XT**	**UM47-NT**	**UM47-XT**
Binary SVM comparisons	**UM22-NT vs UM22-XT**	(6+0+0)/6	(7+0+0)/7	-	-
	UM47-NT vs UM47-XT	-	-	(4+5+0)/9	(8+2+0)/10
	UM22-NT vs UM47-NT	(5+1+0)/6	-	(9+0+0)/9	-
	UM22-XT vs UM47-XT		(7+0+0)/7		(10+0+0)/10

Bibliography

[1] J. Fowler, The rationale of dose fractionation, *The relationship of time and dose in the radiation therapy of cancer* 1969, 6-23.

[2] H. R. Withers, Biologic basis for altered fractionation schemes, *Cancer* **55**(S9), 1985, 2086-2095.

[3] C. Fakhry, W. H. Westra, S. Li, A. Cmelak, J. A. Ridge, H. Pinto, et al., Improved survival of patients with human papillomavirus–positive head and neck squamous cell carcinoma in a prospective clinical trial, *Journal of the National Cancer Institute* **100**(4), 2008, 261-269.

[4] K. Lindel, K. T. Beer, J. Laissue, R. H. Greiner and D. M. Aebersold, Human papillomavirus positive squamous cell carcinoma of the oropharynx: A radiosensitive subgroup of head and neck carcinoma, *Cancer* **92**(4), 2001, 805-813.

[5] D. M. Brizel, R. K. Dodge, R. W. Clough and M. W. Dewhirst, Oxygenation of head and neck cancer: Changes during radiotherapy and impact on treatment outcome, *Radiotherapy and Oncology* **53**(2), 1999, 113-117.

[6] D. Brizel, G. Sibley, L. Prosnitz, R. Scher and M. Dewhirst, Tumor hypoxia adversely affects the prognosis of carcinoma of the head and neck, *International Journal of Radiation Oncology* Biology* Physics* **38**(2), 1997, 285-289.

[7] M. Nordsmark, S. Bentzen, V. Rudat, D. Brizel, E. Lartigau, P. Stadler, et al., Prognostic value of tumor oxygenation in 397 head and neck tumors

after primary radiation therapy. An international multi-center study, *Radiotherapy and Oncology* **77**(1), 2005, 18-24.

[8] M. Baumann, T. Herrmann, R. Koch, W. Matthiessen, S. Appold, B. Wahlers, et al., Final results of the randomized phase iii chartwel-trial (aro 97-1) comparing hyperfractionated-accelerated versus conventionally fractionated radiotherapy in non-small cell lung cancer (nsclc), *Radiotherapy and Oncology* **100**(1), 2011, 76-85.

[9] T. Turrisi, K. Kim, R. Blum, W. T. Sause, R. B. Livingston, R. Komaki, et al., Twice-daily compared with once-daily thoracic radiotherapy in limited small-cell lung cancer treated concurrently with cisplatin and etoposide, *New England Journal of Medicine* **340**(4), 1999, 265-271.

[10] M. Saunders, S. Dische, A. Barrett, A. Harvey, D. Gibson, M. Parmar, et al., Continuous hyperfractionated accelerated radiotherapy (chart) versus conventional radiotherapy in non-small-cell lung cancer: A randomised multicentre trial, *The Lancet* **350**(9072), 1997, 161-165.

[11] Q. Wei, L. Cheng, W. K. Hong and M. R. Spitz, Reduced DNA repair capacity in lung cancer patients, *Cancer research* **56**(18), 1996, 4103-4107.

[12] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, et al., Label-free biomedical imaging with high sensitivity by stimulated raman scattering microscopy, *Science* **322**(5909), 2008, 1857-1861.

[13] P. Matousek and N. Stone, Development of deep subsurface raman spectroscopy for medical diagnosis and disease monitoring, *Chem Soc Rev* **45**(7), 2016, 1794-1802.

[14] C. Matthaus, C. Krafft, B. Dietzek, B. R. Brehm, S. Lorkowski and J. Popp, Noninvasive imaging of intracellular lipid metabolism in macrophages by raman microscopy in combination with stable isotopic labeling, *Analytical chemistry* **84**(20), 2012, 8549-8556.

[15] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, et al., Label-free biomedical imaging with high sensitivity by stimulated raman scattering microscopy, *Science* **322**(5909), 2008, 1857-1861.

[16] S. K. Paidi, A. Rizwan, C. Zheng, M. Cheng, K. Glunde and I. Barman, Label-free raman spectroscopy detects stromal adaptations in premetastatic lungs primed by breast cancer, *Cancer research* **77**(2), 2017, 247-256.

[17] P. T. Winnard Jr, C. Zhang, F. Vesuna, J. W. Kang, J. Garry, R. R. Dasari, et al., Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct raman spectral signatures and metabolomes, *Oncotarget* **8**(12), 2017, 20266.

[18] Barman, N. C. Dingari, A. Saha, S. Mcgee, L. H. Galindo, W. Liu, et al., Application of raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy, *Cancer research* **73**(11), 2013, 3206-3215.

[19] R. Sathyavathi, A. Saha, J. S. Soares, N. Spegazzini, S. Mcgee, R. R. Dasari, et al., Raman spectroscopic sensing of carbonate intercalation in breast microcalcifications at stereotactic biopsy, *Scientific reports* **5**(2015,

[20] S. Holton, M. Walsh, A. Kajdacsy-Balla and R. Bhargava, Label-free characterization of cancer-activated fibroblasts using infrared spectroscopic imaging, *Biophysical journal* **101**(6), 2011, 1513-1521.

[21] T. Kwak, A. Kajdacsy-Balla, V. Macias, M. Walsh, S. Sinha and R. Bhargava, Improving prediction of prostate cancer recurrence using chemical imaging, *Scientific reports* **5**(2015,

[22] E. M. Barroso, R. W. Smits, C. G. Van Lanschot, P. J. Caspers, I. Ten Hove, H. Mast, et al., Water concentration analysis by raman spectroscopy to determine the location of the tumor border in oral cancer surgery, *Cancer research* **76**(20), 2016, 5945-5953.

[23] Q. Matthews, M. Isabelle, S. J. Harder, J. Smazynski, W. Beckham, A. G. Brolo, et al., Radiation-induced glycogen accumulation detected by single cell raman spectroscopy is associated with radioresistance that can be reversed by metformin, *PloS one* **10**(8), 2015, e0135356.

[24] S. J. Harder, Q. Matthews, M. Isabelle, A. G. Brolo, J. J. Lum and A. Jirasek, A raman spectroscopic study of cell response to clinical doses of ionizing radiation, *Applied spectroscopy* **69**(2), 2015, 193-204.

[25] M. S. Vidyasagar, K. Maheedhar, B. M. Vadhiraja, D. J. Fernendes, V. B. Kartha and C. M. Krishna, Prediction of radiotherapy response in cervix

cancer by raman spectroscopy: A pilot study, *Biopolymers* **89**(6), 2008, 530-537.

[26] S. J. Harder, M. Isabelle, L. Devorkin, J. Smazynski, W. Beckham, A. G. Brolo, et al., Raman spectroscopy identifies radiation response in human non-small cell lung cancer xenografts, *Scientific Reports* **6**(2016, 21006.

[27] Alhallak, S. V. Jenkins, D. E. Lee, N. P. Greene, K. P. Quinn, R. J. Griffin, et al., Optical imaging of radiation-induced metabolic changes in radiation-sensitive and resistant cancer cells, *Journal of Biomedical Optics* **22**(6), 2017, 060502-060502.

[28] P. Stein, A. D. Swick, M. A. Smith, G. C. Blitzer, R. Z. Yang, S. Saha, et al., Xenograft assessment of predictive biomarkers for standard head and neck cancer therapies, *Cancer medicine* **4**(5), 2015, 699-712.

[29] S. K. Paidi, S. Siddhanta, R. Strouse, J. B. Mcgivney, C. Larkin and I. Barman, Rapid identification of biotherapeutics with label-free raman spectroscopy, *Anal Chem* **88**(8), 2016, 4361-4368.

[30] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan and A. Gorzsás, Vibrational spectroscopic image analysis of biological material using multivariate curve resolution-alternating least squares (mcr-als), *Nature protocols* **10**(2), 2015, 217.

[31] P.-Y. Sacré, P. Lebrun, P.-F. Chavez, C. D. Bleye, L. Netchacovitch, E. Rozet, et al., A new criterion to assess distributional homogeneity in hyperspectral images of solid pharmaceutical dosage forms, *Analytica Chimica Acta* **818**(2014, 7-14.

[32] W. H. W., Dealing with a common problem in social science: A simplified rank-biserial coefficient of correlation based on the u statistic, *European Journal of Social Psychology* **2**(4), 1972, 463-465.

[33] U. Thissen, B. Üstün, W. J. Melssen and L. M. Buydens, Multivariate calibration with least-squares support vector machines, *Anal Chem* **76**(11), 2004, 3099-3105.

[34] C.-C. Chang and C.-J. Lin, Libsvm: A library for support vector machines, *ACM transactions on intelligent systems and technology (TIST)* **2**(3), 2011, 27.

[35] F. Lyng, E. Gazi and P. Gardner, Preparation of tissues and cells for infrared and raman spectroscopy and imaging, *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach Royal Society of Chemistry, Cambridge* 2010, 147-191.

[36] A. Arnold, J. J. Burmeister and G. W. Small, Phantom glucose calibration models from simulated noninvasive human near-infrared spectra, *Anal Chem* **70**(9), 1998, 1773-1781.

[37] Halberg, T. Khan, M. E. Trujillo, I. Wernstedt-Asterholm, A. D. Attie, S. Sherwani, et al., Hypoxia-inducible factor 1α induces fibrosis and insulin resistance in white adipose tissue, *Molecular and cellular biology* **29**(16), 2009, 4467-4483.

[38] J.-O. Moon, T. P. Welch, F. J. Gonzalez and B. L. Copple, Reduced liver fibrosis in hypoxia-inducible factor-1α-deficient mice, *American Journal of*

Physiology-Gastrointestinal and Liver Physiology **296**(3), 2009, G582-G592.

[39] D. M. Gilkes, S. Bajpai, P. Chaturvedi, D. Wirtz and G. L. Semenza, Hypoxia-inducible factor 1 (hif-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing p4ha1, p4ha2, and plod2 expression in fibroblasts, *Journal of Biological Chemistry* **288**(15), 2013, 10819-10829.

[40] K. H. Hofbauer, B. Gess, C. Lohaus, H. E. Meyer, D. Katschinski and A. Kurtz, Oxygen tension regulates the expression of a group of procollagen hydroxylases, *The FEBS Journal* **270**(22), 2003, 4515-4522.

[41] D. E. Lee, K. Alhallak, S. V. Jenkins, I. Vargas, N. P. Greene, K. P. Quinn, et al., A radiosensitizing inhibitor of hif-1 alters the optical redox state of human lung cancer cells in vitro, *Scientific Reports* **8**(1), 2018, 8815.

[42] T. A. Wynn and L. Barron, Macrophages: Master regulators of inflammation and fibrosis, *Seminars in liver disease* **30**(03), 2010, 245-257.

[43] J. M. Straub, J. New, C. D. Hamilton, C. Lominska, Y. Shnayder and S. M. Thomas, Radiation-induced fibrosis: Mechanisms and implications for therapy, *Journal of Cancer Research and Clinical Oncology* **141**(11), 2015, 1985-1994.

[44] E. Rysman, K. Brusselmans, K. Scheys, L. Timmermans, R. Derua, S. Munck, et al., De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation, *Cancer research* **70**(20), 2010, 8117-8126.

[45] J. Mims, N. Bansal, M. S. Bharadwaj, X. Chen, A. J. Molina, A. W. Tsang, et al., Energy metabolism in a matched model of radiation resistance for head and neck squamous cell cancer, *Radiation research* **183**(3), 2015, 291-304.

[46] Y.-C. Kao, S.-W. Lee, L.-C. Lin, L.-T. Chen, C.-H. Hsing, H.-P. Hsu, et al., Fatty acid synthase overexpression confers an independent prognosticator and associates with radiation resistance in nasopharyngeal carcinoma, *Tumor Biology* **34**(2), 2013, 759-768.

[47] T. Shimura, Acquired radioresistance of cancer and the akt/gsk3β/cyclin d1 overexpression cycle, *Journal of Radiation Research* **52**(5), 2011, 539-544.

[48] P. M. Diaz, S. V. Jenkins, K. Alhallak, D. Semeniak, R. J. Griffin, R. P. M. Dings, et al., Quantitative diffuse reflectance spectroscopy of short-term changes in tumor oxygenation after radiation in a matched model of radiation resistance, *Biomedical Optics Express* **9**(8), 2018, 3794-3804.

[49] Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, et al., Intraoperative brain cancer detection with raman spectroscopy in humans, *Science translational medicine* **7**(274), 2015, 274ra219-274ra219.

[50] J.-X. Cheng and X. S. Xie, Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine, *Science* **350**(6264), 2015, aaa8870.

Chapter 6

Thesis summary and future directions

6.1 Thesis summary

The studies presented in this thesis show that label-free Raman spectroscopy can take advantage of its synergistic combinations with advancements in nanotechnology and the growing number of multivariate data analysis tools, including but not limited to machine learning, for quantitative molecular characterization of cancer.

Our SHINERS investigation reported in Chapter 2 revealed that while the enhancement of signal resulted in lower acquisition time, higher accuracy of prediction using the PLS-DA derived detection algorithm and better contrast between closely related pathologies, spontaneous Raman spectroscopy also provided encouraging results across the classes studied. By separately analyzing the data from the sites that harbored microcalcifications, we found that these samples provided higher classification accuracies using SHINERS and spontaneous Raman in comparison to those that didn’t harbor microcalcifications. Overall, the studies in this chapter showed that SHINERS can provide a route for rapid label-free disease diagnosis and subtyping in clinical samples or margin assessment during tissue conserving surgeries.

This chapter is partially adapted from the published encyclopedia article: Paidi SK, Pandey R and Barman I. “Medical applications of Raman spectroscopy”, Encyclopedia of Analytical Chemistry, 1-21, 2020.

The studies in Chapter 3 helped identify Osteopontin (OPN) as a key driver of the microcalcification and metastasis in breast cancer. By stable shRNA knockdown of the OPN gene in the aggressive MBA-MB-231 breast cancer cells and alizarin red S staining for microcalcifications, we showed a significant reduction in the stain intensity and OPN mRNA expression levels in the OPN silenced clones in comparison to the empty-vector control cells. The OPN silenced clones also exhibited a significant reduction in their in vitro migration potential. Furthermore, these observations were supported by elevated OPN levels in the circulating tumor cells and lung metastatic cells in vivo. Taken together, our studies in this chapter hinted at a dual role of OPN in directing microcalcification formation and cell migration.

In Chapter 4, we showed that Raman spectroscopy can identify pre-metastatic changes in the lung of mice bearing breast cancer xenografts. We used principal component analysis to identify spectral markers indicative of changes in collagen and proteoglycan content of the ECM in pre-metastatic lungs and noted that the changes scale with metastatic potential of the primary tumor. These observations were corroborated using histology and gene expression analysis. Finally, our classification routine based on partial least squares-discriminant analysis allowed the identification of an inadequately primed animal. Overall, our studies showed that Raman spectroscopy can be a valuable addition to the arsenal of current tools used for studying the composition and evolution of pre-metastatic niches in secondary organs.

Finally, in Chapter 5, we extend our Raman investigations to study the response of tumors to radiation therapy and predict resistance to therapy. We employed sensitive and resistant variants of isogenic lung cancer cells and head and neck cancer cells to grow tumors in athymic nude mice and subjected them to X-ray therapy at clinically relevant fractionated doses and sham radiation. By subjecting the Raman spectra obtained from irradiated and unirradiated tumors to multivariate curve resolution-alternating least squares, we found significant model-specific compositional differences in the tumor microenvironment between irradiated and unirradiated samples for both sensitive and resistant tumors. Additionally, we built classifiers based on support vector machine (SVM) to assess the utility of Raman spectra data for prediction of response and resistance to radiation therapy.

Overall, our studies pave the way for establishing Raman spectroscopy as a tool for both clinical decision-making in cancer care and unraveling molecular mechanisms underlying disease progression.

6.2 Future work

To date, several proof-of-principle studies using Raman spectroscopy have been conducted to understand a range of diseases and their progression both in vivo and ex vivo. However, the widespread clinical translation of this attractive technology requires further improvements in the following areas. First, inclusion of larger patient cohort for each application is critical for development of robust classification models that can effectively capture inter-patient variability. As a next step in this direction, standardizing protocols for acquisition of Raman spectral datasets from clinical samples will help compare the results obtained by various groups and potentially merge them into larger datasets. Second, translation of SERS technologies in vivo will require additional tests to confirm their toxicity and accumulation in human body. However, such studies are challenging due to use of a variety of synthesis protocols, reporter molecules, capping agents, buffers, coatings and surface functionalization by different research groups. Thirdly, uniform metrics of success such as accuracy, sensitivity, specificity and statistical power must be defined by clinicians to compare and select spectroscopic protocols that are suitable for clinical translation. To accomplish this step, Raman spectroscopy and other emerging optical techniques must be introduced in medical education to promote awareness and enthusiasm among clinicians about this emerging suite of techniques. Such efforts are expected to promote more active collaborations between clinicians and spectroscopists.

In addition to clinical diagnostics, Raman spectroscopy and its variants (SERS in particular) are rapidly emerging as important tools for discovery of molecular markers that characterize differences between different diseased states along novel axes. The preliminary spectral markers obtained from analysis of Raman datasets can potentially yield a narrow subset of biomolecules to be employed as targets for future studies using conventional wet chemistry approaches. Similarly, studies that have attempted to combine Raman spectroscopy with other optical spectroscopy techniques such as autofluorescence, diffuse reflectance spectroscopy and optical coherence tomography have shown promise for symbiotic enrichment of its capabilities.

Vita

Santosh Kumar Paidi graduated from the Indian Institute of Technology Bombay in 2014 with a Bachelor of Technology in Mechanical Engineering and a minor in Aerospace Engineering before commencing doctoral study at Johns Hopkins University. His research efforts in Dr. Ishan Barman's lab were directed towards the application of Raman spectroscopy and multivariate data analysis to develop novel quantitative approaches for addressing unmet needs in the molecular study of cancers. Overall, his doctoral research has resulted in 15 peer-reviewed publications in journals such as Cancer Research, Nano Letters and Analytical Chemistry. He has been awarded the Tomas A. Hirschfeld Scholar Award, the Coblentz Student Award, the Barbara Stull Graduate Student Award, the SLAS Graduate Student Fellowship Grant, Whiting School Doctoral Fellowship, Molecular Medicine Tri-Conference Student Fellowship and Undergraduate Research Award by IIT Bombay in recognition of his work. In roles such as GRO Advocacy Chair and WSE Representative on the Homewood Graduate Board, Santosh strove for enriching graduate student experience at JHU. Outside Hopkins, he volunteers for outreach programs aimed at encouraging the involvement of school students in STEM fields. Santosh currently serves as an associate editor for the Journal of Emerging Investigators, which publishes original research conducted by middle and high school students.