Raman spectroscopy of isogenic breast cancer cells derived from organ-specific metastases reveals distinct biochemical signatures

by
Chi Zhang

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland

May, 2015

Abstract

Characterizing the subtle divergence of the metastatic lesions from the primary tumor is critical to understanding organ-specific adaptations that regulate further disease progression as well as the development of targeted chemotherapy treatment options. Though genomic assays have provided insights into the aberrant expression of a few biomarkers, dissecting metastatic cancers based on objective molecular markers still remains challenging. I show that the exquisite specificity of Raman microspectroscopy in detecting molecular phenotypes can be harnessed to investigate and differentially identify engineered metastatic breast cancer cellular models in a label-free manner. A Raman microscope is used to acquire spectra from unique organ-specific human metastatic breast cancer cell lines that were established from the outgrowth of metastatic breast cancer cells from explant cultures of each organ. By correlating the Raman spectra with the pathology, I architected partial least squares-discriminant analysis and support vector machine-derived decision algorithms that exhibit significant power in segmenting between the established cell lines in the brain, lung, liver, spine and breast. Using the acquired chemical profiles, I show the robustness of the method to spurious correlations and ascertain the informative spectral bands that hint at organ-specific biomarkers as opposed to the presence of a single universal marker. These findings underscore the significance of tissue-specific microenvironments, especially the lipid phenotype in promoting adaptations in metastatic cancer cells and highlight the potential of Raman spectroscopy for further evaluation of targeted chemotherapeutic approaches in these cellular model systems.

Advisor: Prof. Ishan Barman

Acknowledgments

These researches cooperate with Associate Professor Venu Raman and Paul T. Winnard Jr. of Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine.

Also these researches received support from Jeon Woong Kang of Laser Biomedical Research Center, Massachusetts Institute of Technology.

The relative paper is under reviewing by Analytical Chemistry.

Contents

List of Tables

List of Figures

Introduction

Breast cancer is the most common malignant neoplasm and is the second leading cause of cancer-related death among women in the United States, exceeded only by lung cancer [1]. Recent advances in the understanding of breast cancer progression, increased mammographic screening and the development of novel therapeutic modalities have positively impacted mortality rates with the American Cancer Society recently reporting a 5-year survival rate near 99% for local breast cancer [2]. However, the 5-year survival for metastatic breast cancer that involves distant organs drops to a dismal 24% [2]. Our understanding of metastatic breast cancer is still rudimentary, resulting in our limited ability to accurately predict and monitor the condition. Critically, obtaining safe chemotherapeutic regimen strategies that ablate metastatic lesions is an unmet clinical need with the current practice of systemic administration of cytotoxic chemotherapy having very limited effect on survival, which results in numerous adverse side-effects and no cures [3].

When considering solutions to this problem an important factor is the divergence of the metastatic cancer cells growing in an organ outside of the breast from the primary breast tumor. This is evident as cancer cell populations are characteristically heterogeneous displaying various degrees of genomic instability as well as dynamic adaptations to survive fluctuating microenvironmental conditions. The organ-specificity of the metastatic spread needs to be a critical consideration, as clinical treatment decision options for distant metastatic breast cancer have historically relied, in part, on an evaluation of a few select biomarkers found during assessment of the primary tumor [4]. In fact, recent evidence from retrospective and prospective clinical trials indicates that

matched primary breast tumor and metastatic lesion biopsy samples often exhibit divergent expression of markers such as ER and HER-2 [5]. To explain these findings, researchers have hypothesized that metastasis organotropism emerges through acquisition of distinct sets of organ-specific metastasis genes in metastatic variants that are best adapted to different target organ microenvironments through Darwinian selection. Consequently, only those cancer cells that become imbued with traits that favor survival in each organ will thrive and impair organ function [6]. Due to varying from a same primary cell line, it is very difficult to discern various organ-specific metastatic lesions to the prescribed therapeutic regimen. Accordingly, the organ-specificity of the metastatic spread needs critical reconsideration, as historically clinical treatment decision options for distant metastatic breast cancer have relied, in part, on an evaluation of a few select biomarkers found during assessment of the primary tumor, such as ER and Her-2. This strategy although beneficial to some extent, is also a likely contributing factor to the diminished response rates for survival from metastatic disease as metastatic lesions might present with altered biomarker signatures than the corresponding primary tumor [4, 7]. In addition, the present bank of molecular profiles of matched primary and metastatic breast tumors do not facilitate patient specific smart therapeutic alternatives.

Dissecting metastatic cancers based on objective molecular markers still remains challenging. Indeed, many clinical studies have correlated alterations in expression of individual genes with breast cancer disease outcome with contradictory results [8]. Here, I propose a fundamentally different approach towards identification of metastatic cancer cells and selection of relevant molecules involved in the metastatic spread. Harnessing the exquisite specificity of Raman microspectroscopy in detecting molecular phenotypes

in cells and tissue, I aimed to obtain rapid and label-free profiling of newly generated isogenic metastatic human breast cancer cellular models. Given its lack of sample preparation requirements and ability to provide quantitative biochemical analyses in near real-time conditions, Raman spectroscopy provides a powerful tool for live cell analysis. While this spectroscopic technique has been recently used to distinguish between malignant, normal, and benign breast tissues, by us and others [9-12], the potential for using these spectral markers as new routes to recognition of metastatic cell types that are isogenic to the primary tumor, has been surprisingly underappreciated.

In this study, a Raman microscope is used to record spectra from unique organ-specific human metastatic breast cancer cell lines, which were established from the outgrowth of metastatic breast cancer cells from explant cultures of brain, liver, lung, and spine as well as the primary orthotopic xenograft site, *i.e.*, mammary fat pad (MFP) tumors (Fig. 1).

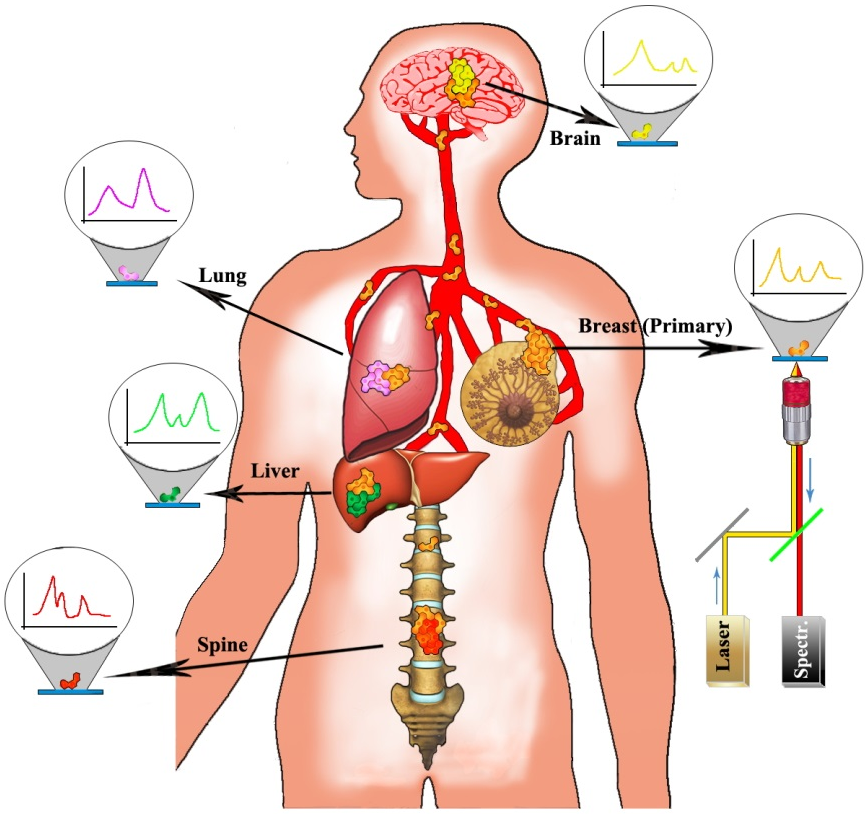

Figure 1 Schematic illustration of the Raman spectroscopy measurements of organ-specific metastatic breast cancer cell lines. The pathways and distant metastases sites to which the primary tumor cells (orange) migrate are shown here. The breast cancer cells that adapt and thrive in the brain microenvironment are shown in yellow. Similarly, the metastatic cellular deposits in the lung, liver and spine are represented in pink, green and red. A Raman microspectroscope is used to record spectra from these organ-specific breast tumor cell lines that are established from the outgrowth of metastatic breast cancer cells from explant cultures of each organ.

I used Raman spectroscopic measurements reveal the presence of subtle, but consistent, spectral differences of the cell lines. Using multivariate chemometric methods, I show that these spectral changes emanating from the variations of specific biochemical attributes at each metastatic site can be utilized to architect decision algorithms with high diagnostic power. Specifically, I report that partial least squares discriminant analysis (PLS-DA)-based decision algorithm can offer an average correct classification rate of ~95% in discerning between the signatures of metastatic cell lines grown out of brain,

lung, liver, spine, and MFP. Furthermore, I identify the presence of spectrally informative features that bring to light putative unique biomarkers for each site probably as a result of the intricate tumor-stroma interactions at the target organ. Importantly, these findings underscore the relevance of Raman spectral information in characterizing isogenic metastatic lesions at different sites in terms of inherent biochemical determinants and that this can be accomplished in a non-destructive manner without staining or requiring *a priori* knowledge of the molecular transformations.

Materials and methods

Mice

All animal handling procedures were performed in accordance with protocols approved by the Johns Hopkins University Institutional Animal Care and Use Committee and conformed to the Guide for the Care and Use of Laboratory Animals published by the NIH. Non-Diabetic severe combined immunodefcient (NOD-SCID) female mice, ages 6 to 8 weeks, were used throughout these studies. At the end of the experiments, mice were sacrificed by administering an overdose of anesthetic [saline:ketamine:acepromazine (2:1:1)] followed by cervical dislocation.

Cell Culture and Treatments

All MDA-MB-231-tdTomato (231-tdT) culturing was done in standard humidified incubators at 37^{o} C and 5% CO_2. Primary tumors were initiated by injection of $2x10^6$ 231-tdT cells into the second thoracic mammary fat pad of 5 female NOD-SCID mice. After 13 - 15 weeks of growth when primary tumors were on average 1200 cm^3 the mice were sacrificed and primary tumor, brain, liver, lungs and spine, were immediately excised, dissected away from fat and muscle, placed into sterile PBS on ice. Pieces of primary tumor, heavily diseased lungs and small portion of liver with a macroscopic metastatic lesion were then immediately minced in 100 mm cell culture dishes containing 10 ml of medium within a sterile hood. All other organs/bones were inspected using fluorescence microscopy for any signs of metastatic burden, which was easily discerned as bright tdT red fluorescence. Areas of fluorescence along with adjacent tissue were cut away and placed into cell culture plates in sterile medium.

All organ/bone tissue explants were initially cultured in RPMI-10% FBS supplemented with antibiotics (100 I.U./ml penicillin, 100 µg/ml streptomycin, 100 µg/ml ampicillin, and 100 µg/ml kanamycin) and, as necessary, Fungizone. The latter was often used during culturing cells out of spine as these pieces of bone were large, tended to float, i.e, became collagen rafts, and thus somewhat exposed at the medium to air surface, which promoted fungal growth. Medium was refreshed every 2-3 days and after two weeks of culture the medium was changed to RPMI-10% FBS supplemented with pen/strep. During routine passages the medium/floating cells was first collected and the adherent colonies were then lifted of the plates by room temperature incubations in HANKS-5 mM EDTA solution for 2–5 min with shaking and tapping by hand. Lifted cells were pooled with the collected medium/cells, centrifuged 200xg at 21^{o} C for 10 mins, and the supernatant (medium-EDTA) discarded. Cell pellets were then suspended in fresh medium and plated at the desired densities. It took at least 24 hours and often 48 hours (generally during recovery from -80^{o} C storage) for the larger percentage of adherent cells to settle and start to grow.

Growth Rate

Growth curves (Fig. 2) were generated by seeding 24 well plates with 10^5 cells per well and harvesting quadruplicates of these wells every 24 hours through to the 144 hours end-point. Live cell counts were obtained with a TC10 Automatic Cell Counter (Bio-Rad) in the presence of Trypan Blue. Despite being isogenic, these cell lines exhibit important growth distinctions that support my hypothesis that each metastatic site imbues metastatic tumors with unique specific molecular attributes that differ from site-to-site.

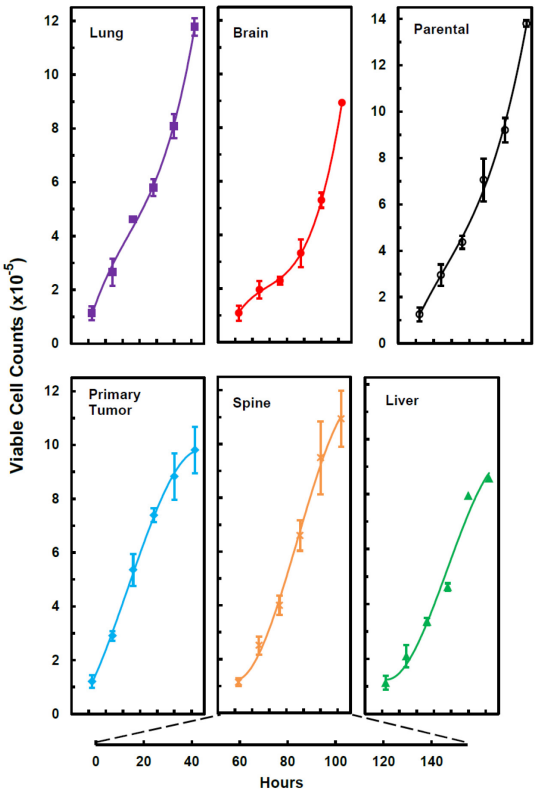

Figure 2 Growth curves used to estimate growth rates as doubling times during log-phase growth. Qualitatively, the growth curves of lung and brain cell lines have a similar shape and appear not to have reached a stationary phase of growth. Correspondingly, primary tumor, liver, and spine cell line growth curves exhibited similar S-shapes that end in stationary phases of growth.

Motility Assay

Standard motility assays were done in 24 well Transwell® plates (Costar) with 8.0 µm membrane inserts. Cells were seeded onto duplicate wells at a density of 2000 cells/well. Three separate experiments were done using RPMI-5% FBS medium in the lower chamber while the inserts with cells contained RPMI-0.2% FBS medium. At the end of the 2 week culturing time, colony numbers at the bottom surface of membranes were counted using the inherent red fluorescence of tdT as a "stain" with a fluorescence microscope (Nikon Eclipse TS100) and 4x objective. For each experiment two fields of view were counted from each well.

Optical Microscopy

Phase contrast and fluorescence microscopy was done on a Nikon ECLIPSE TS 100 microscope (Nikon Instruments, Inc.) equipped with a Photometrics CoolSnap ES digital camera (Roper Scientific), and FITC and Texas Red filter cubes. The fluorescence light source was an X-Cite 120 Fluorescence Illumination System (Photonic Solutions, Inc.). Images were collected with NIS-Elements F3.2 software and processed with ImageJ.

Raman Microspectroscopy

The custom-built Raman microscope used in this work was previously reported [13, 14]. A 785 nm Ti: Sapphire laser (3900S, Spectra-Physics), pumped by a frequency-doubled solid-state laser (Millennia 5sJ, Spectra-Physics), was used as the excitation source for the inverted microscope (Fig. 3).

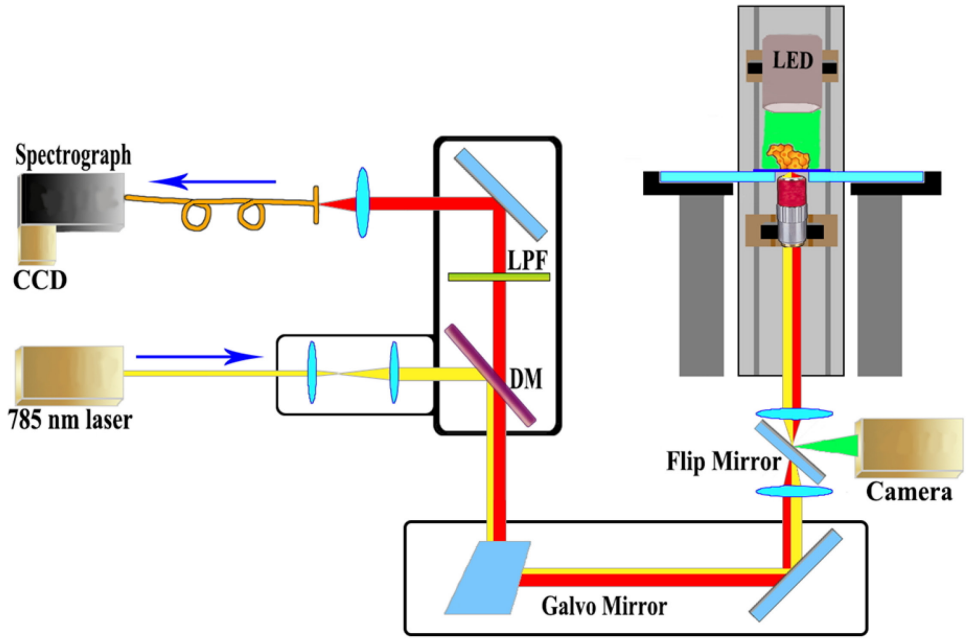

Figure 3 High-throughput Raman microspectroscopy system. The system incorporates confocal Raman, confocal reflectance (not shown here) and bright field imaging modalities for visualization and characterization of unstained live cells. LPF: Long Pass Filter; DM: Dichroic Mirror.

The laser was focused onto the specimen using a 1.2 NA water immersion objective lens (UPLSAPO60XWIR 60X, Olympus) that also functioned to collect the backscattered signal. The collected signal was then recorded using a TE-cooled, deep depletion CCD (1340/400-EB, Princeton Instruments) following dispersion through an imaging spectrograph (HoloSpec f/1.8i, Kaiser Optical Systems). Additionally, bright field and phase contrast microscopy was performed for visualization and registration with the Raman measurements. Instead of interrogating single cells at the subcellular level, the ultimate goal of the current study is to characterize biochemical variances at the ensemble cellular level, and thus a collection of cells in pellets were investigated using point spectroscopic measurements. After replacing culture medium with PBS, cell pellets were formed by centrifugation and placed on top of the quartz coverslip for Raman measurement. 100 (10 × 10) spectra were collected from 90μm × 90μm area in each pellet with axial resolution of 25 μm. Raman spectra were recorded by vertical binning before averaging of 10 successive frames, each with an acquisition time of 0.3 sec, for a total collection time of 3 sec. Wavelength calibration was performed prior to spectral acquisition by acquiring spectra from 4-acetamidophenol, a Raman scatterer with well-characterized peak positions. The 600-1800 cm^{-1} fingerprint region was used for the ensuing analysis (spectral resolution of 8 cm^{-1}). Cosmic ray removal was also implemented before the spectra were subjected to multivariate statistical analysis in MATLAB (Mathworks Inc.).

Multivariate Statistical Analysis

While Raman microspectroscopy provides a promising tool, in principle, to non-invasively probe biological specimen with high specificity, its intrinsic weak signals

(especially in relation to conventional fluorescence imaging) and spectral complexity provides a substantive challenge in univariate or ratiometric quantitation of the sample constituents. Hence, to arrive at biochemical variances in isogenic cellular sublines, multivariate statistical analysis was performed on the acquired Raman spectra. By exploiting the full spectral information, as opposed to focusing on a single peak, multivariate techniques provide a robust route in extracting information both amenable and hidden from human examination.

In this study, the Raman spectra were first subjected to principal component analysis (PCA). PCA is a widely used exploratory data analysis technique and employs dimension reduction to amplify the subtle differences in the recorded spectral profiles [15]. Operating without any *a priori* knowledge of the samples, PCA seeks to determine an alternate set of linearly uncorrelated coordinates, *i.e.* principal components (PC), such that the maximum variance in the spectral data can be explained by using only a few PCs. In particular, I employed PC scores to reveal the clustering behavior – or the lack thereof – between the metastatic breast cancer cell sublines, and the coefficient loadings to uncover the critical diagnostic variables/regions in the spectra associated with the underlying differences in the spectral data.

Additionally, to develop decision algorithms for predicting the cell type (class membership) of the spectra, partial least squares-discriminant analysis (PLS-DA) and support vector machines (SVM) were used. The former employs PLS analysis for noise reduction and variable selection and determines the maximal separation between each class by fitting a unique global model to the entire dataset. The number of loading vectors incorporated in the decision algorithm is determined by the leave-one-out cross validation

procedure (LOOCV) [16]. Similar to PLS-DA in its supervised nature, SVM is rooted in statistical learning theory and structural risk minimization concepts and designs separating boundaries between classes by solving a constrained quadratic optimization problem. I used a radial basis function (RBF) with a Gaussian envelope to enable the separation of classes in a higher dimensional space and the optimization and kernel parameters were determined based on an automated grid search algorithm. Two different classification methods were used to confirm the validity of the results and to minimize the possibility of spurious correlations that may plague an "overfitted" decision algorithm. The output of the PLS-DA and SVM-derived decision algorithms was validated against the known class labels, *i.e.* the specific line of the metastatic breast cancer cellular model system. The performance of the algorithms was evaluated by determining the sensitivity and specificity using a LOOCV protocol. Similar approaches to classification of Raman spectroscopic data have been described elsewhere in the literature [17, 18].

Results

In order to facilitate the tracking of metastatic progression in live mice, triple negative MDA-MB-213 human breast cancer cells were engineered to be a cell line which stably express a red fluorescence protein (231-tdT). Thus, due to the inherent very bright fluorescence of 231-tdT cells, this mouse model provides facile *ex vivo* fluorescence microscopic identification of metastatic lesions within any organ of choice. As shown in Fig. 4, although the organ explants are resolved as only amorphous material (bright field images in Fig. 4) without visually discernable metastatic lesions, the red fluorescence revealed the presence of the cancer.

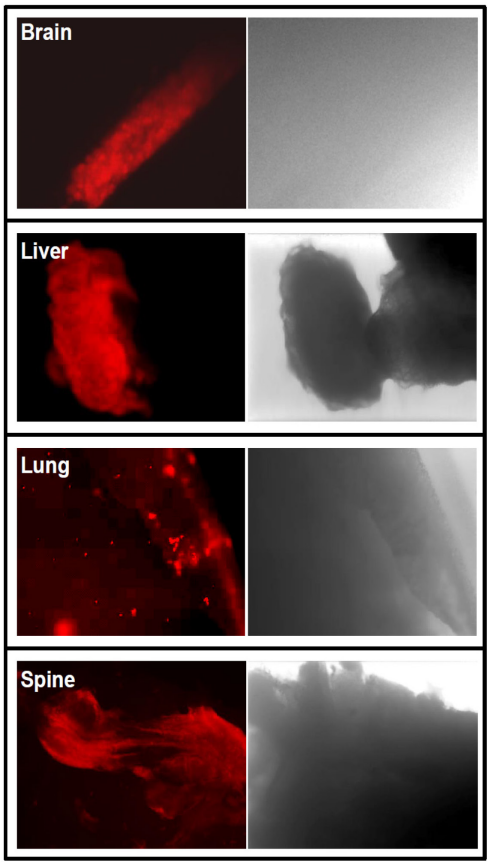

Figure 4 Fluorescence images of metastatic 231-tdT lesions in fresh organ samples. Fluorescence images of brain, liver, lung, and spine tissues immediately after dissection.

Raman Spectroscopic Differentiation of Organ-specific Metastatic Isogenic Breast Cancer Cell Lines.

The mean Raman spectra of raw data with ±1 standard deviations (SD) of the metastatic isogenic breast cancer cell lines corresponding to the primary tumor from the orthotopic MFP site, brain, liver, lung, and spine are shown in Fig. 5 (the spectra are normalized and offset for visualization purposes). The latter four sites are representative of the common clinically observed breast cancer metastatic destinations [19].

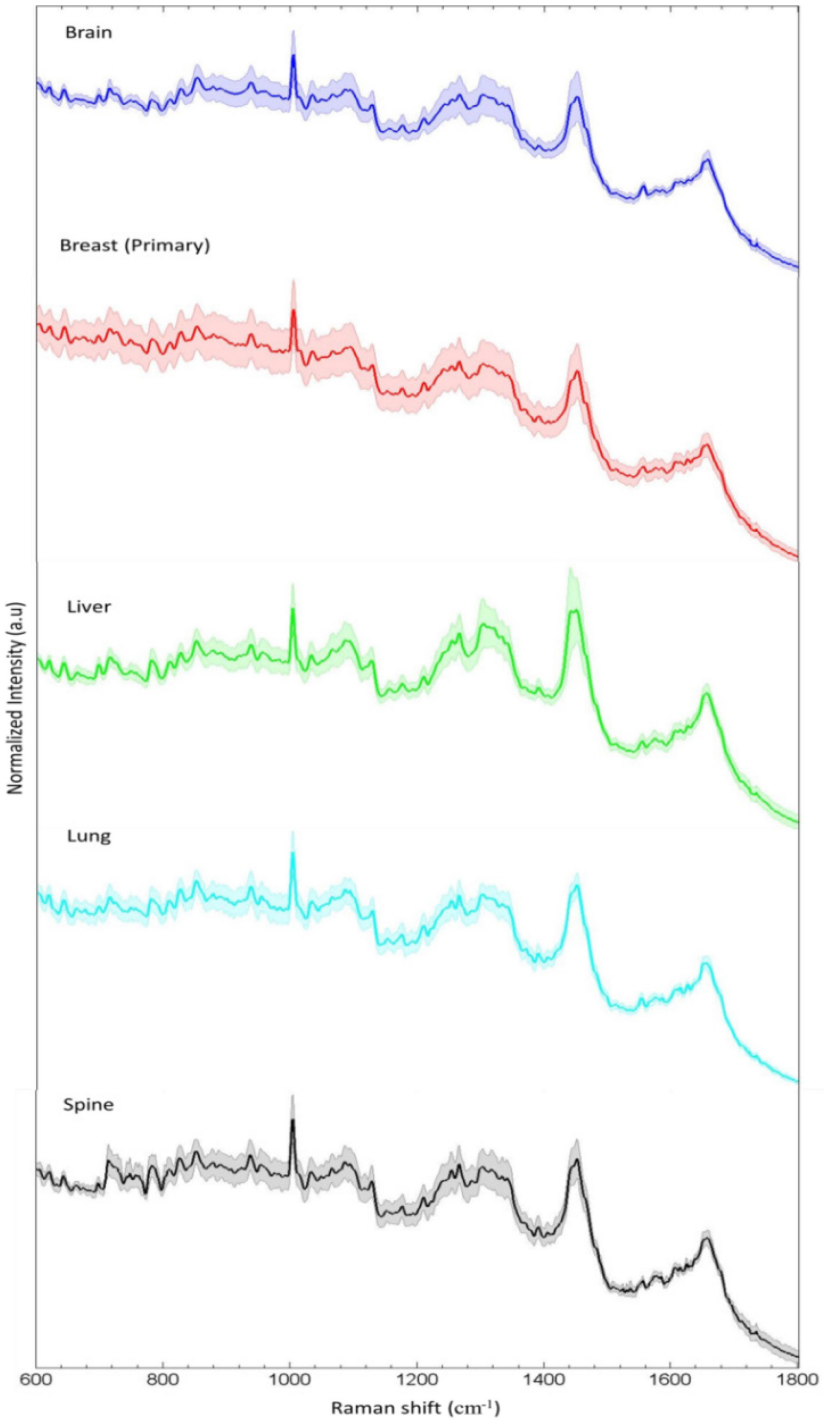

Figure 5 Representative Raman spectra of organ-specific metastatic breast cancer cell lines. Spectra are acquired from the brain (blue), breast (red), liver (green), lung (cyan) and spine (black) cell lines. The solid profile depicts the mean spectrum of each sample group and the shadow represents ±1 standard deviation. Spectra are normalized and offset for visualization.

Distinctive Raman peaks located at around 852, 937, 1005, 1090, 1265, 1305, 1334, 1452 and 1657 cm^{-1} are seen for all the cell lines with lower intensity features observable at 782, 878 and 1067 cm^{-1}, respectively. In agreement with previous reports [20], the features seen here can primarily be attributed to the different vibrational modes of proteins (852, 878, 937, 1005, 1265, 1305, 1335, 1452 and 1657 cm^{-1}), lipids (1305 and 1452 cm^{-1}) and nucleic acids (782, 1067, 1090 and 1335 cm^{-1}). Though the spectra grossly appear to have similar profiles, careful inspection reveals subtle but discernible and reproducible shape differences, especially on removal of the fluorescence background [21]. I reasoned that while the subtle differences in the spectral dataset and small variations within the profiles of each cell line impede the possibility of differentiation using single-feature analysis, multivariate classification methods could enable recognition and segmentation of the cell pathology provided the between-class distinctions are reproducible and surpass within-class differences.

To examine the tumor cell lines and ascertain the differential biochemical characteristics that define each cell line, I employed principal component analysis (PCA) to transform the dimensions of the acquired spectral profiles into an alternate set of linearly uncorrelated variables (*i.e.* principal components, PC), along which the variation in the data is maximal (Fig. 6).

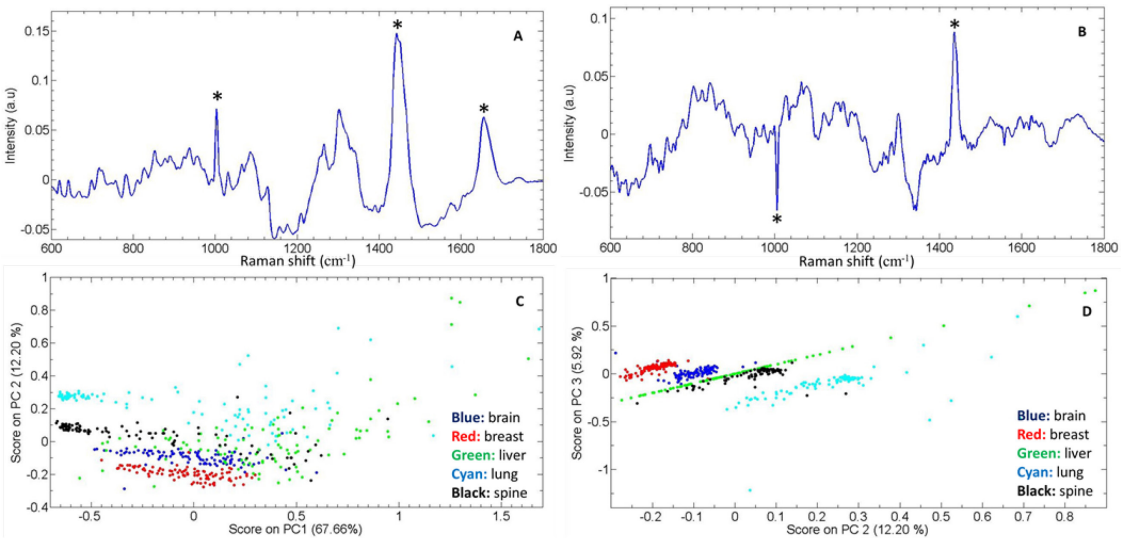

Figure 6 Principal components loadings and scores plot for the Raman measurements from all the cell lines. (A) & (B) show the loadings for principal component (PC) 1 and 2, respectively. The prominent peaks in each case are highlighted by asterisks, including features at 1005 cm^{-1}, 1452 cm^{-1} and 1657 cm^{-1}. (C) & (D) provides an illustration of the scores from the cell lines corresponding to PC1, PC2 and PC3. Percentages in the score plots represent the variance accounted for by each PC. Blue: brain, Red: breast, Green: liver, Cyan: lung, Black: spine.

This dimensional reduction step is critical to enabling sample exploration via visual assessment of similarities and differences between samples and, ultimately, in identifying the smallest possible subset of discriminatory features necessary to build a robust decision algorithm. PC1 and PC2 scores accounted for approximately 67% and 12% of the total variance in the dataset. I observe that PC1 prominently features the Raman scattering peaks of 1005 cm^{-1} (ν_s(C-C) symmetric ring breathing of phenylalanine), 1452 cm^{-1} (ν (C-N) in-plane vibration) and 1657 cm^{-1} (ν (C=O) of amide I of proteins) (Fig. 6*A*). As such, the PC1 loading consists of features that collectively indicate its representation of the protein spectral profile. Significantly, the PC2 profile exhibits a negatively directed phenylalanine feature with positively directed in-plane vibration features common to both proteins and lipids (Fig. 6*B*). Monitoring a subset of these

spectral markers provides important clues to defining the metastatic cancer cell lines in molecular terms, i.e., by correlating the differential molecular expression to the organ-specificity of the cell line. I postulate that the PC2 loading provides a measure of the lipid content and that the lipid and protein cell content are inversely correlated, based on the juxtaposition of the negatively directed phenylalanine feature (a common marker for proteins) with the positively directed feature at 1452 cm^{-1} (common to both lipids and proteins). This has direct implications for the corresponding PC scores plot (Fig. 6*C* & *D*). The scores plots reveal substantive (though not perfect) clustering of the metastatic cell lines using only three variables. The spread over a larger area of the PC scores axes indicates the possible presence of more heterogeneity in the lung and liver cell lines particularly in relation to the brain and primary tumor cell lines. Importantly, based on my hypothesis, I interpret that the primary tumor and brain cell lines have the lowest lipid content (i.e., PC2 score). In contrast, the spine, liver and lung cell lines exhibit higher levels of lipid concentrations and clear separation from the other cell lines. Prior reports of the organ-specific pattern of breast tumor metastasis in which the bone (60%), lung (34%) and liver (20%) are the organs most commonly affected lend support to my PCA-based discrimination of these three cell lines [22-25]. My observation also hints at an underlying relationship between exacerbated lipogenesis, metastatic potential and organ-specificity of breast cancer cells.

To quantify the segmentation capability using the spectroscopic measurements, I developed decision algorithms based on partial least squares discriminant analysis (PLS-DA) and support vector machines (SVM). The number of loading vectors (LV) used in the PLS-DA model was determined based on the minimal misclassification rate in a

leave-one-out cross-validation protocol while ensuring that the spectra to LV ratio was greater than 5 to avoid problems of data sparseness. Subsequently, the dataset was split into training (70% of the spectra) and test (30%) sets to estimate the classification accuracy. This entire operation: re-splitting, training of the decision algorithm, and prediction, was performed 1000 times to obtain outcomes with well-defined statistical confidence (Table 1). The overall classification accuracy obtained for the PLS-DA-derived decision algorithm was found to be 96.8% with the classification accuracy for each cell line being in excess of 93%. The SVM-derived decision algorithm also provides similar levels of classification performance (Table 2) affirming that the richness of the spectral data is the principal driver for the prediction performance.

Table 1 Classification outcomes in prospective prediction for the PLSDA-derived decision algorithm

Average correct classification rate: 96.8%

Reference Diagnosis	Correct Classification	Misclassification
Brain	98.0 %	2.0 %
Primary Tumor	99.3 %	0.7 %
Liver	97.4 %	2.6 %
Lung	93.3 %	6.7 %
Spine	96.1 %	3.9 %

Table 2 Classification outcomes in prospective prediction for the SVM-derived decision algorithm

Average correct classification rate: 97.6%

Reference diagnosis	Correct classification rate	Misclassification rate
Brain	99.6 %	0.4 %
Primary Tumor	98.9 %	1.1 %
Liver	94.3 %	5.7 %
Lung	97.3 %	2.7 %
Spine	98.1 %	1.9 %

Additionally, to ensure the robustness of these findings, I implemented a negative control study. In this case, the labels (primary tumor, brain, liver, lung and spine) were assigned in a randomized order, regardless of their actual identity. Using the acquired spectra in conjunction with these control labels, I re-derived the PLS-DA and SVM decision algorithms and used them in the same analysis protocol as detailed previously. In this situation, a low correct classification rate for each cell line was obtained with the average rate of correct classification below 20% (approximately the random chance of predicting correctly 1 class out of a total of 5 classes). This underscores the robustness of the spectroscopic measurements to confounding variables and chance correlations. Additionally, the derived decision algorithms should not be impacted by systematic temporal correlations since the reported experiments were conducted over several days in a randomized manner. Taken together, these results demonstrate that Raman

spectroscopy offers a reliable tool for discriminating these isogenic metastatic breast cancer cell lines on the basis of distinct organ-of-origin driven biochemical adaptations.

Discussion

My findings provide strong evidence that Raman spectroscopic signatures can be used to investigate molecular differences between breast cancer cells from diverse metastatic sites by probing the biochemical phenotypic variances. While the Raman spectra provided a label-free, quantitative measure of the specimen's molecular composition, the stochastically varying intracellular compositions and complex spatial distributions of the molecules precluded single feature evaluation. Thus, I used multivariate statistical analysis to yield decision algorithms that are robust with respect to stochastic variance and offer real-time segmentation capabilities. These algorithms exploit subtle differences in the vibrational signatures of the molecular markers that are reflective of the multiple and complex interactions between metastatic cells and host homeostatic mechanisms.

To clarify the segmentation capability, I sought further specificity biochemical differences in cellular model system. I performed difference analysis across the normalized spectra obtained from pairwise comparisons of cell lines to demarcate the informative regions with the goal of identifying biomarkers, which would be either universal or characteristic to the specific pair of cell lines. Fig, 7 shows the comparison between primary tumor and liver cell lines. Fig. 7A black spectrum represents the liver cell lines data and red spectrum shown in Fig. 7B is the data from primary tumor. Fig. 7C is the differences between primary tumor and liver cell lines, only the relative differences' absolute values larger than 0.01 can be counted as a considerable variation from primary tumor. The potential components assignments for each variation are listed in Table 3.

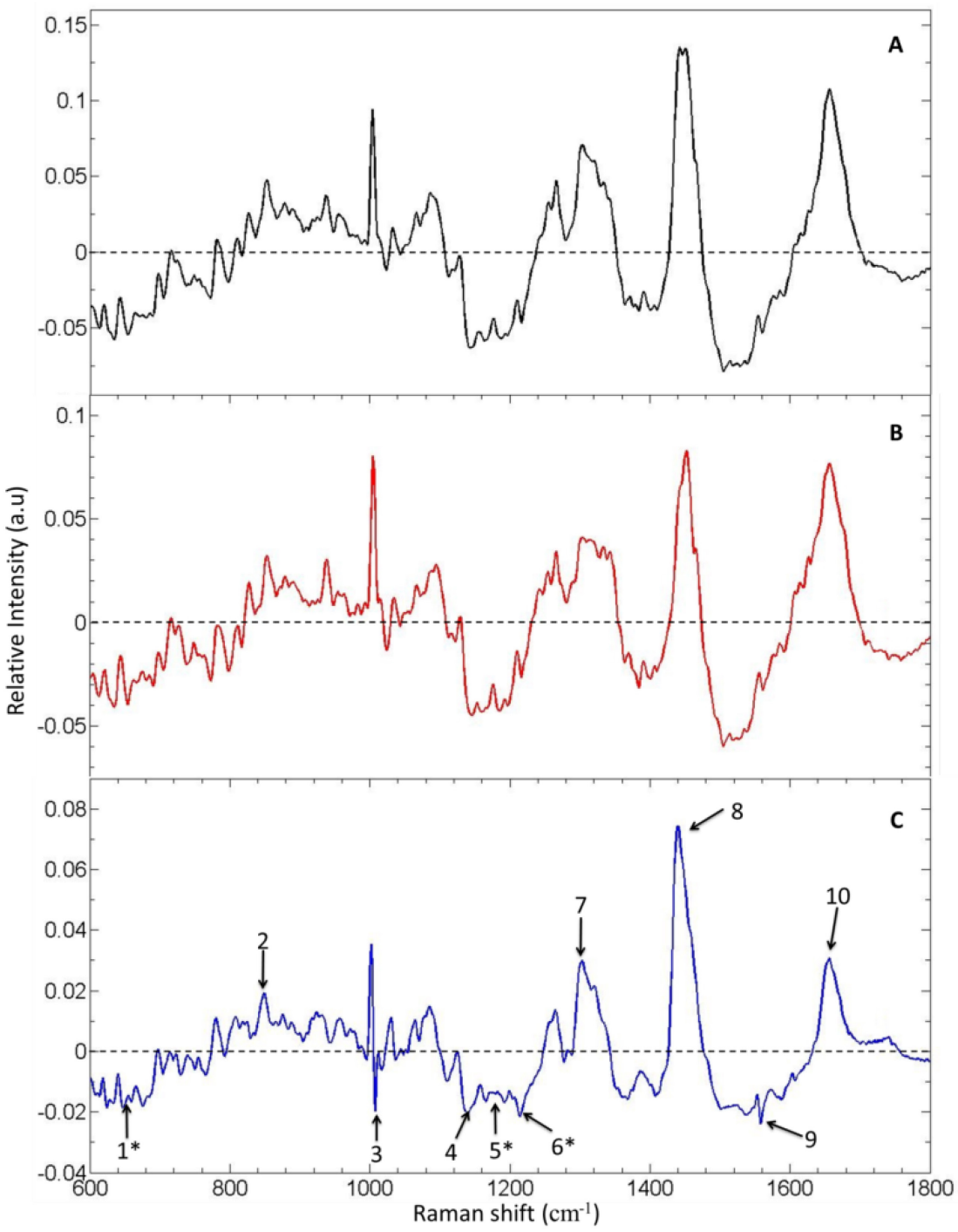

Figure 7 Liver Raman spectrum and primary tumor Raman spectrum comparison. (A) & (B) show the average liver cell lines spectrum and average primary tumor spectrum. (C) is the differences between liver and primary tumor on Raman spectra. The considerable variants comparing with primary tumor are marked in (C).

Table 3 Differences between Breast (primary) and Liver whose absolute values exceed criterion

No.	Range	Exact Raman shift	Component assignment	Comparison with primary
1*	603-689 cm^{-1}	621 cm^{-1}	C-C twist in phenylalanine	negative
1*		645 cm^{-1}	C-C twist in tyrosine	
1*		671 cm^{-1}	C-S stretching in cysteine	
2	840-854 cm^{-1}	854 cm^{-1}	Ring breathing in tyrosine/C-C stretching in proline, polysaccharides	positive
3	998-1009 cm^{-1}	1006 cm^{-1}	Symmetric ring breathing mode of phenylalanine	positive
4	1130-1233 cm^{-1}	1129 cm^{-1}	Skeletal C-C stretching in lipids	negative
5*		1160 cm^{-1}	C-C/C-N stretching in protein	
5*		1180 cm^{-1}	Cytosine/guanine/adenine	
6*		1220 cm^{-1}	Amide III: β-sheet	
6*	1258-1268 cm^{-1}	1258 cm^{-1}	Amide III: β-sheet/adenine/cytosine, CH2 in-plane deformation (lipids)	positive
7	1294-1330 cm^{-1}	1308 cm^{-1}	CH2 deformation in lipids/adenine/cytosine	positive
8	1429-1470 cm^{-1}	1452 cm^{-1}	CH2 deformation in lipids, fatty	positive

			acids	
9	1488-1598 cm^{-1}	1582 cm^{-1}	Adenine/guanine, Amide II, tryptophane	negative
10	1643-1677 cm^{-1}	1661 cm^{-1}	Amide I: α-helix	positive

Fig. 8, 9 & 10 are the comparisons between primary tumor and spine, or lung, or brain. All of them follow the same protocol of comparison of primary tumor and liver. Among of them, brain cell lines show the smallest variants. Table 4, 5 & 6 are the components assignments corresponding to comparisons of spine, lung and brain.

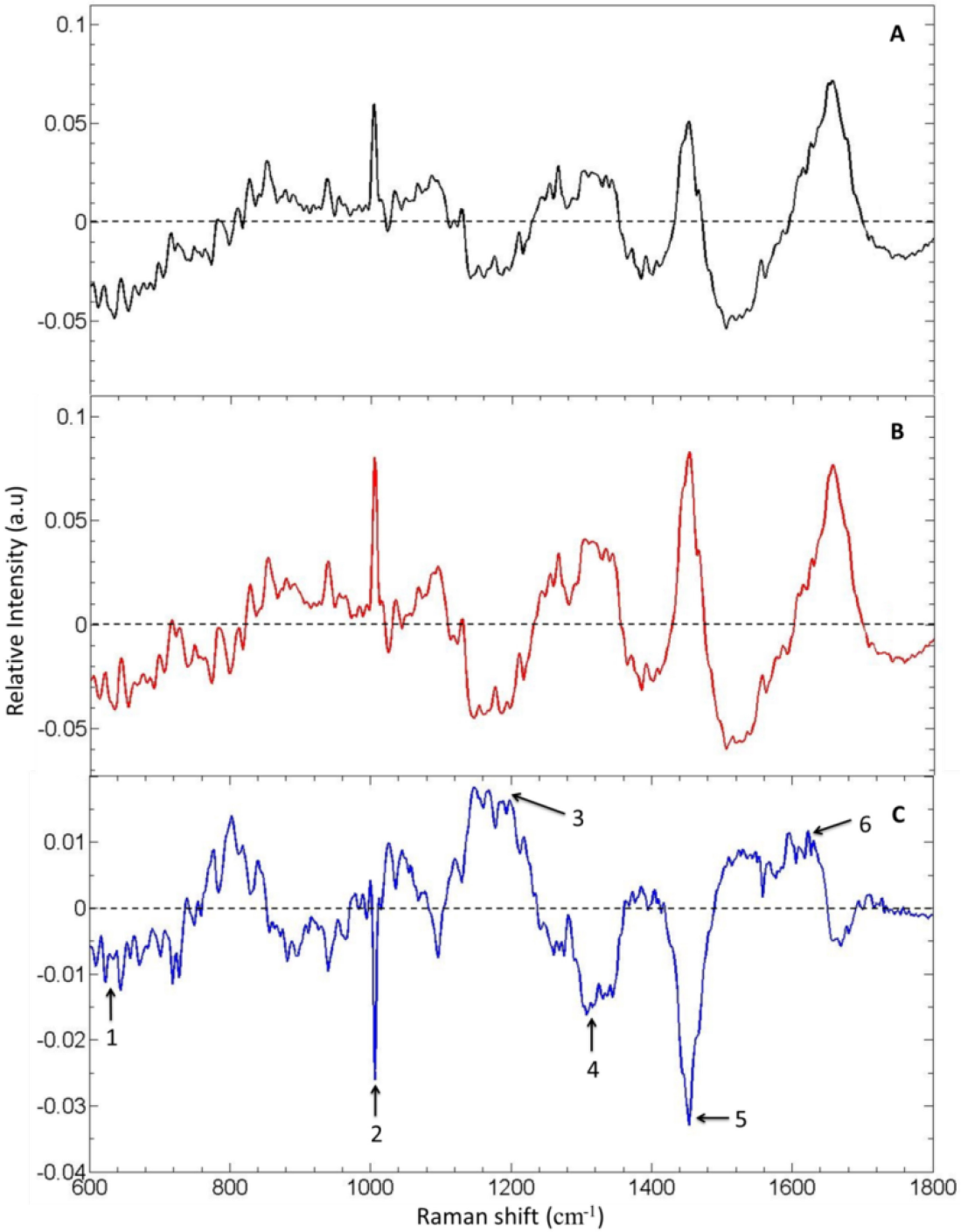

Figure 8 Spine Raman spectrum and primary tumor Raman spectrum comparison. (A) & (B) show the average spine cell lines spectrum and average primary tumor spectrum. (C) is the differences between spine and primary tumor on Raman spectra. The considerable variants comparing with primary tumor are marked in (C).

Table 4 Differences between Breast (primary) and Spine whose absolute values exceed criterion

No.	Range	Exact Raman shift	Component assignment	Comparison with primary
1	623 cm^{-1}	623 cm^{-1}	C-C twist in phenylalanine	negative
2	1006 cm^{-1}	1006 cm^{-1}	Phenylalanine NADH, Symmetric ring breathing mode of phenylalanine	negative
3	1136-1211cm^{-1}	1160 cm^{-1}	C-C/C-N stretching (proteins)	positive
4	1298-1350 cm^{-1}	1335 cm^{-1}	CH3/CH2 twisting or bending mode of lipids/collagens, CH2 deformation in lipids/adenine/cytosine	negative
5	1435-1472 cm^{-1}	1442 cm^{-1}	Fatty acids, CH2 (lipids and proteins), CH2 deformation in lipids	negative
6	1622 cm^{-1}	1622 cm^{-1}	Tryptophan	positive

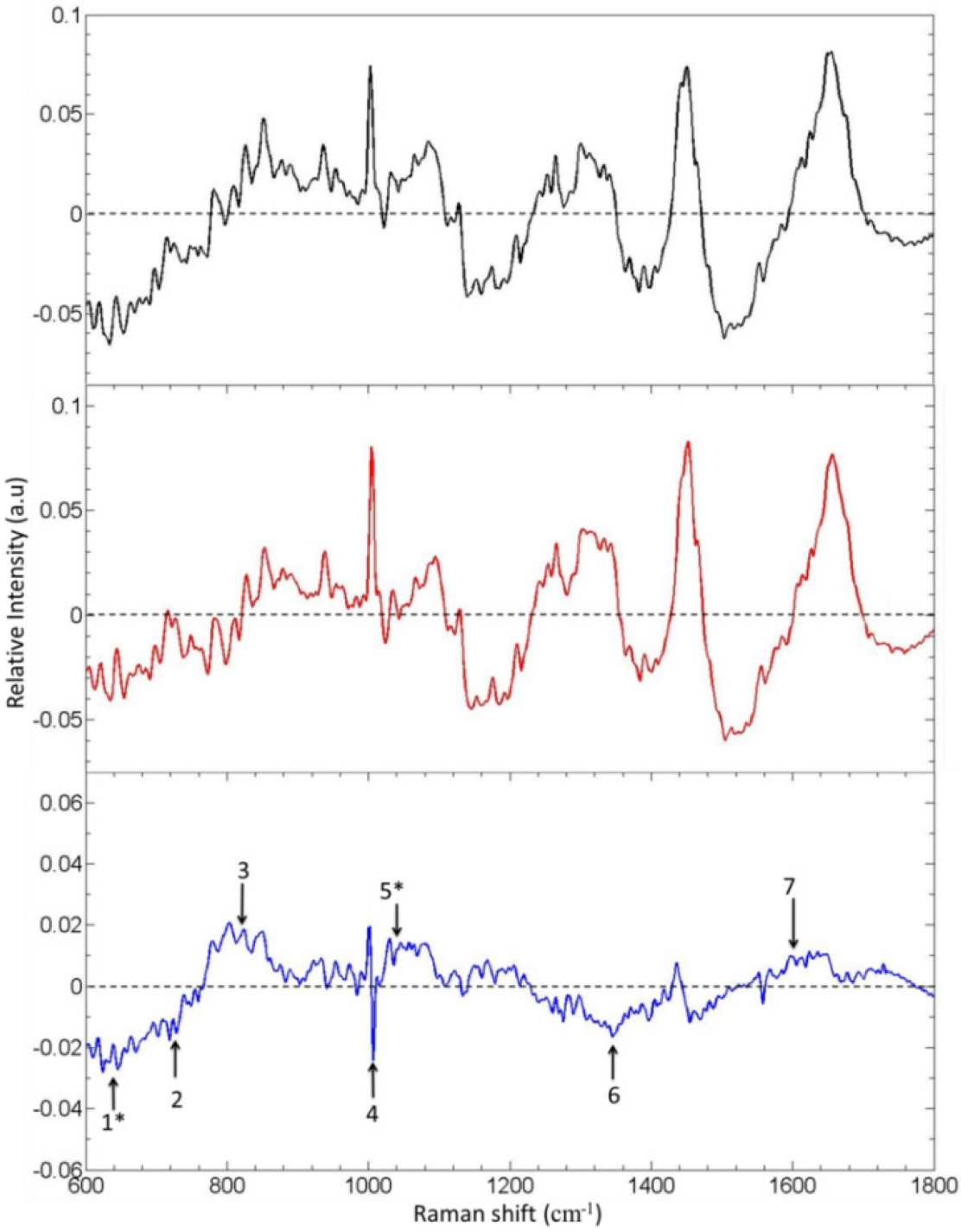

Figure 9 Lung Raman spectrum and primary tumor Raman spectrum comparison. (A) & (B) show the average lung cell lines spectrum and average primary tumor spectrum. (C) is the differences between lung and primary tumor on Raman spectra. The considerable variants comparing with primary tumor are marked in (C).

Table 5 Differences between Breast (primary) and Lung whose absolute values exceed criterion

No.	Range	Exact Raman shift	Component assignment	Comparison with primary
1*	512-733 cm^{-1}	621 cm^{-1}	C-C twist in phenylalanine	negative
1*		645 cm^{-1}	C-C twist in tyrosine	
2	775-856 cm^{-1}	788 cm^{-1}	DNA: O-P-O backbone stretching /thymine/cytosine, proline	positive
3		833 cm^{-1}	DNA: O-P-O backbone stretching/out of plane ring breathing in tyrosine, polysaccharides	
4	1000-1006 cm^{-1}	1006 cm^{-1}	Symmetric ring breathing mode of phenylalanine	positive
5*	1026-1089 cm^{-1}	1036 cm^{-1}	C-H in plane bending mode of phenylalanine, proline	positive
5*		1071 cm^{-1}	Skeletal C-C stretch in lipids	
6	1311-1361 cm^{-1}	1340 cm^{-1}	Polynucleotide chain (DNA bases), CH3CH2 wagging collagen, nucleic acid	negative
7	1623 cm^{-1}	1623 cm^{-1}	Tryptophan	positive

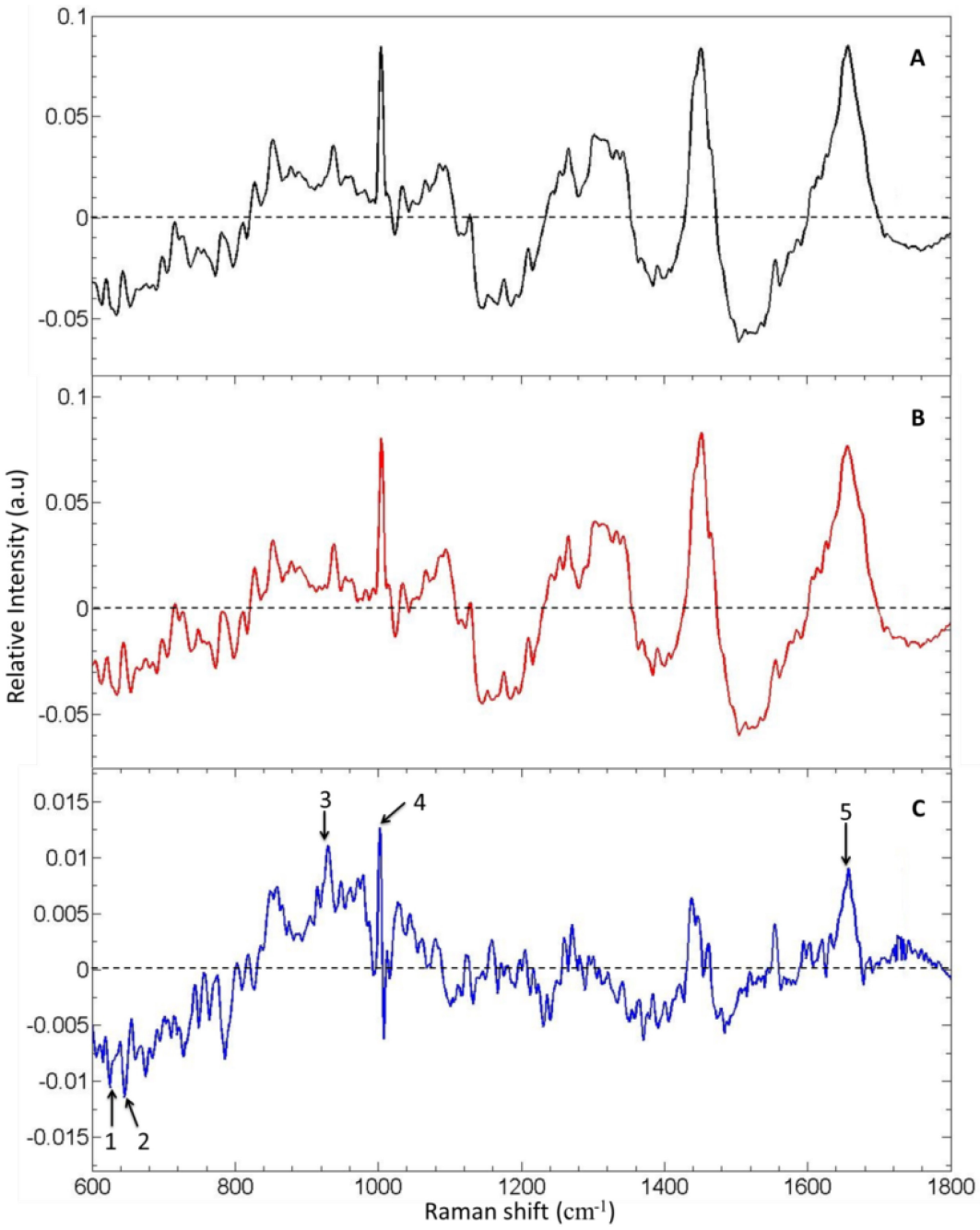

Figure 10 Brain Raman spectrum and primary tumor Raman spectrum comparison. (A) & (B) show the average brain cell lines spectrum and average primary tumor spectrum. (C) is the differences between brain and primary tumor on Raman spectra. The considerable variants comparing with primary tumor are marked in (C).

Table 6 Differences between Breast (primary) and Brain whose absolute values exceed criterion

No.	Range	Exact Raman shift	Component assignment	Comparison with primary
1	625 cm^{-1}	625 cm^{-1}	C-C twist in phenylalanine	negative
2	645 cm^{-1}	645 cm^{-1}	C-C twist in tyrosine	negative
3	931 cm^{-1}	931 cm^{-1}	C-C skeletal stretching in protein, proline ring/glucose/lactic acid	positive
4	1002-1008 cm^{-1}	1006 cm^{-1}	Symmetric ring breathing mode of phenylalanine	positive
5	1658 cm^{-1}	1658 cm^{-1}	Amide I: α-helix, Amide I, Lipid	positive

Scan through these considerable variants regions to find out common wavenumber ranges across all five organ-specific breast tumors, several common regions were pointed out. Using the primary tumor and liver cell lines as a representative case, I observed that the prominent features in the difference spectra, i.e., above noise level, are due to intensity differences at the 1305 and 1452 cm^{-1} peaks, both of which are common to lipids and proteins and a first derivative-like feature at 1005 cm^{-1} phenylalanine band. Unpaired two-sided Student's *t*-tests also reveal that Raman peak intensities in these regions are significantly different as also in the 1136-1211 cm^{-1} region. PCA classification on these two cell lines also highlights the importance of these regions (Fig. 11A). Furthermore, the PC2 obtained here displays similarly directed features at both the

lipid markers (1305 and 1452 cm^{-1}) and oppositely directed peaks at protein-only markers; e.g., 1005 cm^{-1}.

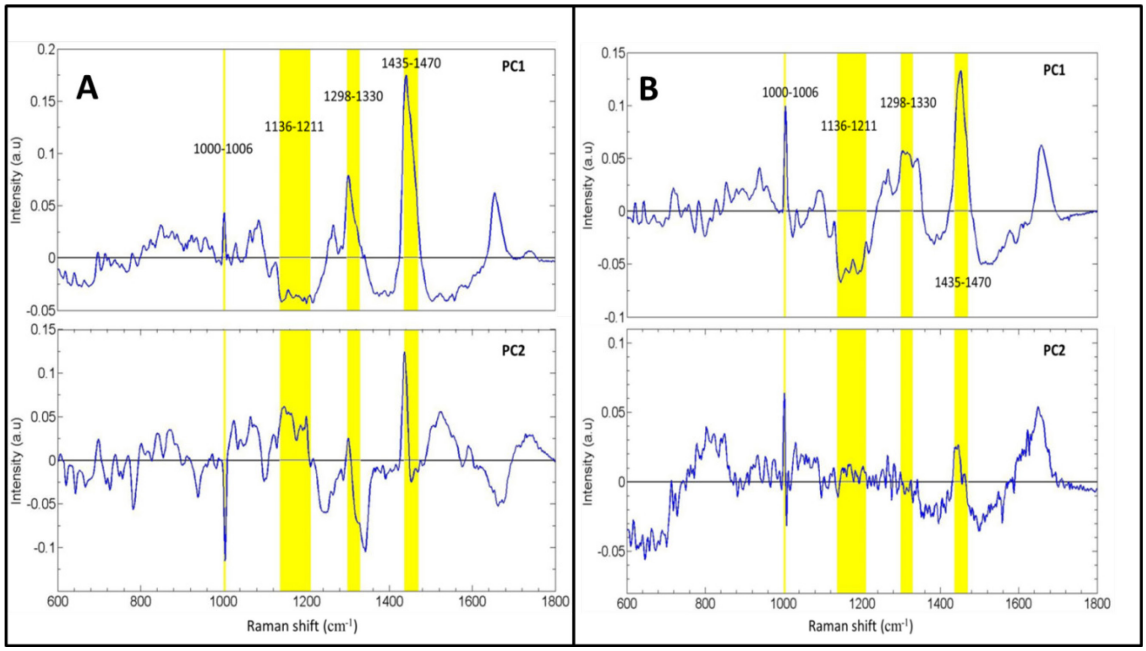

Figure 11 Identification of informative spectral regions via PCA data exploration of liver cell lines, spine cell lines and primary tumor. (A) and (B) are PCA outcomes of liver cell lines, primary tumor and spine cell lines, primary tumor, respectively. Illustration of the PC loadings corresponding to the spectral dataset acquired from primary breast cancer and liver metastasized tumor cell lines. The top and bottom panels show PC1 and PC2 loadings, respectively. The highlighted bars (yellow) represent the wavelength regions elucidated from the difference spectra and unpaired Student's t-test as those with the most significant variability from the cell lines. Evidently, these regions also include the critical features constituting the PC spectral profiles.

The PCA analyses across other pairs of cell lines validate these findings from comparison of the primary and liver cell lines, for example, PCA of primary breast tumor and spine cell line is provided in Fig. 11B.

Using only the selected regions (highlighted by the yellow bars of Fig. 11A & B), I developed a PLSDA-derived decision algorithm to reclassify all the cell lines that provided equally impressive prediction performance (Table 7) as that obtained using the full spectral analysis. In this case, only 9.6% of the spectral information was used

indicating that the model is based on the molecular-specific information elucidated from the cell lines. The feature selection process, thus, not only avoids the "*curse of dimensionality*" (and the possibility of creating spurious models) but also enables the development of a robust classifier that is based on a few, biologically relevant and interpretable discriminatory features [26].

Table 7 Classification outcomes in prospective prediction for the PLSDA-derived decision algorithm using only the biomarker-specific wavelength bands

Average correct classification rate: 91.3 %

Reference diagnosis	Correct Classification	Misclassification
Brain	91.7 %	8.3 %
Primary Tumor	97.2 %	2.8 %
Liver	91.1 %	8.9 %
Lung	85.8 %	14.2 %
Spine	90.6 %	9.4 %

Based on these results, I infer that the lipid content of the metastatic cancer cells provides an indirect measure of a multitude of functions closely linked with the ability to specifically adapt to growth in a variety of tissue microenvironments. As such, this could aid in differentiating between an underlying genetic or metabolic signature of these cell lines that will potentially provide new probes for the identification of the phenotypes. The significance of the lipid content in my isogenic breast cancer cell lines is consistent with

emerging data from other laboratories that have; e.g., observed a correlation between activation of *de novo* lipogenesis and metastatic potential [27, 28, 29]. In particular, it has been shown that elements regulating the pathways for fatty acid synthesis, choline and ethanolamine phospholipid production, as well as cholesterol metabolism, can be closely linked to the metastatic potential. Thus, alterations in lipogenesis processes along with the types of lipids made and utilized could be part of the determinant adaptations that define the specificity metastatic growth in an organ.

Conclusion

In the present study, I have demonstrated that the molecular characterization capability of Raman microspectroscopy, coupled with multivariate statistical analysis, offers a powerful label-free and nondestructive technique for discerning the organ-specific phenotypes of the metastatic breast cancer cells. These important differences point at the fact that organs differ vastly with unique attributes of metabolism, developmental programs, microenvironments, and function resulting defined identities. My findings indicate that Raman microspectroscopy will not only provide significant information towards classification, diagnosis and prognosis of breast cancers but also aid in improving our understanding of mechanisms of breast cancer metastasis. In particular, the investigation of the lipid phenotype using Raman spectroscopic imaging may provide important clues in identification of critical organ-specific determinants controlling colonization and growth and addressing fundamental issues such as the fraction of cells with metastatic potential. Ultimately, I envision these studies will be the key to understanding why present therapies have a minimal impact on controlling metastatic disease and, thus, provide the basis for development of targeted chemotherapeutic approaches in patients with the goal of alleviating pain and prolonging life.

References

1. U.S. Cancer Statistics Working Group (2014) United States Cancer Statistics: 1999–2011 Incidence and Mortality Web-based Report. Available at: www.cdc.gov/uscs.

2. DeSantis CE, et al. (2014) Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin* 64(4):252–271.

3. Ding L, et al. (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 464:999–1005.

4. Aurilio G, et al. (2014) A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. Eur J Cancer 50(2):277-289.

5. Klein-Goldberg A, Maman S, Witz IP (2013) The role played by the microenvironment in site-specific metastasis. Cancer Lett 352(1):54-58.

6. Fidler IJ (2003) The pathogenesis of cancer metastasis: The 'seed and soil' hypothesis revisited. *Nat Rev Cancer* 3(6):453–458.

7. Botteri E, et al. (2012) Biopsy of liver metastasis for women with breast cancer: impact on survival. Breast 21(3):284-288.

8. van 't Veer LJ, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530-536.

9. Haka AS, et al. (2002) Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. *Cancer Res* 62(18):5375–5380.

10. Haka AS, et al. (2005) Diagnosing breast cancer by using Raman spectroscopy. *Proc Natl Acad Sci U S A* 102(35):12371–12376.

11. Stone N, Matousek P (2008) Advanced transmission Raman spectroscopy: A promising tool for breast disease diagnosis. *Cancer Res* 68(11):4424–4430.

12. Barman I, et al. (2013) Application of Raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsies. *Cancer Res* 73(11):3206-3215.

13. Kang JW, et al. (2011) Combined confocal Raman and quantitative phase microscopy system for biomedical diagnosis. *Biomedical Optics Express* 2(9):2484-2492.

14. Kang JW, Nguyen FT, Lue N, Dasari RR, Heller DA (2012) Measuring Uptake Dynamics of Multiple Identifiable Carbon Nanotube Species via High-Speed Confocal Raman Imaging of Live Cells. *Nano Letters* 2(12):6170-6174.

15. Ringnér M (2008) What is principal component analysis. *Nature Biotechnology* 26:303-304.

16. Brereton RG (2003) Chemometrics: Data Analysis for the Laboratory and Chemical Plant. *Lab Automation & Chemometrics*, (John Wiley & Sons, Ltd, Chichester, UK), pp 489.

17. Lim L, et al. (2014) Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis. *J Biomed Opt* 19(11):117003.

18. Soares JS, et al. (2013) Diagnostic power of diffuse reflectance spectroscopy for targeted detection of breast lesions with microcalcifications. *Proc Natl Acad Sci U S A* 110(2):471–476.

19. Zlotnik A, Burkhardt AM, Homey B (2011) Homeostatic chemokine receptors and organ-specific metastasis. *Nat Rev Immunol* 11(9):597-606.

20. Huang Z, et al. (2010) In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy. *Biosens Bioelectron* 26(2):383-389.

21. Lieber CA, Mahadevan-Jansen A (2003) Automated method for subtraction of fluorescence from biological Raman spectra. *Appl Spectrosc* 57(11):1363-1367.

22. Lu X, Kang Y (2007) Organotropism of breast cancer metastasis. *J Mammary Gland Biol Neoplasia* 12(2-3):153–162.

23. Gabriel N, Hortobagyi MD (2000) Developments in chemotherapy of breast cancer. *Cancer* 88(12):3073–3079.

24. Kaal EC, Niël CG, Vecht CJ (2005) Therapeutic management of brain metastasis. *Lancet Neurol* 4(5):289–298.

25. Minn AJ, et al. (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436(7050):518-524.

26. Reddy RK, Bhargava R (2010) Chemometric Methods for Biomedical Raman Spectroscopy and Imaging. *Emerging Raman Applications and Techniques in Biomedical and Pharmaceutical Fields,* eds Matousek P, Morris MD(Springer, Berlin Heidelberg), pp 179-213.

27. Hilvo M, et al. (2011) Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res* 71(9):3236–3245.

28. Bhalla K, et al. (2011) PGC1α promotes tumor growth by inducing gene expression programs supporting Lipogenesis. *Cancer Res* 71(21):6888–6898.

29. Nieva C, et al. (2012) The Lipid Phenotype of Breast Cancer Cells Characterized by Raman Microspectroscopy: Towards a Stratification of Malignancy. *PLoS One* 7(10):e46456.

Curriculum Vit:

Chi Zhang

3010 Guildford Ave., BALTIMORE, MD 21218

Tel: (410)917-8546

Email: czhang55@jhu.edu

EDUCATION AND TRAINING

09/2013-06/2015 Johns Hopkins University

Major: Mechanical Engineering

Degree: Master of Mechanical Engineering

Speciality: Photonics for Quantitative Biology and Medicine

09/2009-07/2013 Beihang University (BUAA), China

Major: Mechanical Design and Automation

Degree: Bachelor of Engineering

Thesis: Structure of Halbach Magnet Array and PMSLM Propulsive Force Analysis (Advisor: Rongying Huang)

PROFESSIONAL EXPERIENCE

01/2013-05/2015 Graduate Research Assistant (Master student)

09/2015-present Graduate Research Assistant (Ph.D. student)

RESEARCH AND INTERNSHIP EXPERIENCES

01/2013-05/2015 Photonics for Quantitative Biology and Medicien Laboratory, Department of Mechanical Engineering, Advisor: Ishan Barman, Johns Hopkins University

* Participated into research and laboratory set up. Took charge of multiple projects' spectra acquisition and data processing. Took charge of Raman spectroscope system components design, purchasing, assembling and software programming by MATLAB and LabVIEW.

07/2012-09/2012 Advanced Robotics Laboratory, Department of Mechanical and Automation Engineering, Faculty of Engineering, Advisor: Yangshen Xu, the Chinese University of Hong Kong

* Collaborated and designed intelligent Shoes System. Took charge of the design and framework of hardware, data processing and analysis with MATLAB and LabVIEW.
* Attended a part of Design of manipulator, rescue robot, and climbing robot on a space station with using Pro/E and SolidWorks; Simulated of the control system and built basic client interface with using MATLAB's SIMULINK.

06/2012--07/2012 Chengdu Aircraft Industry (Group) Co., Ltd. (CAC) Contents: Practiced in lathes, CNC, etc.

* Leaded practice Intern group, participated in the programming of Computer Numerical Control Machine Tools with PID programming and G code.

12/2011-08/2012 Beihang University

* Participated the Dual-duty Roller Skating project in FENGRU CUP Competition.

* Took charge of the design, assembly, improvement of the products and approval defense.

12/2010 -11/2011 Beihang University

* Participated the Automatic Sorting and Sequencer of Micro Switch Component project in Student Research Training Program (SRTP).
* Presided the research group. Managed all the concerned matters, including project present, oral conclusion defense, application, design, personnel management, manufacture, etc.

07/2010--09/2010 Sunline International Business Ltd.

* Provided information and technical supports for the Jilin Tianyuan Petrochemical Co., Ltd.'s Oil Project in Madagascar
* Finished the outline of the Business financing plan for Beijing Sunline Biotechnology

PUBLICATION, PATENT AND CONFERENCE

PUBLICATIONS:

* Paul T. Winnard Jr., **Chi Zhang**, Jeon Woong Kang, Farhad Vesuna, Jonah Grray, Ramachandra Rao Dasari, Ishan Barman, Venu Raman. Raman spectroscopy of isogenic breast cancer cells derived from organ-specific metastases reveals distinct biochemical signatures. *Analytical Chemistry*. (Under reviewing)
* Abigail S. Haka, Erika Sue, Priya Bhardwaj, **Chi Zhang**, Joshua Sterling, Cassidy Carpenter, Madeline Leonard, Maryem Manzoor, Jose Aleman, Daniel Gareau, Peter Holt, Jan L. Breslow, Xi Kathy Zhou, Dilip Giri, Monica Morrow, Neil Iyengar, Ishan

Barman, Clifford A. Hudis, Andrew J. Dannenberg. Raman Spectroscopy for Biopsy-Free Detection of White Adipose Tissue Inflammation. *PNAS*. (Under reviewing)

* [Nicolas Spegazzini, Ashwin Kumar Myakalwar], **Chi Zhang**, Siva Kumar Anubham, Ramachandra R. Dasari, Ishan Barman, Manoj Kumar Gundawar. Less is more: Avoiding the LIBS dimensionality curse through judicious feature selection for explosive detection. *Scientific Reports* (Under reviewing)

* **Chi Zhang**, Yibo Wang, Zhixiang Song, Tengda Liu, Zhuo Li, and Xiaoli Li. Research on an Automatic Directional Sorting Device for Micro-switch Elements. *Research and Exploration in Laboratory*. 2013, 32(4):40-42 and 95.

PATENT:

Chi Zhang, Yibo Wang, Zhixiang Song, Tengda Liu, Zhuo Li and Xiaoli Li. *The Device of Automatic Sorting and Sequencer of Micro Switch.* ZL201120493584.2, 08/01/2012

CONFERENCE:

* **Chi Zhang**, Paul Winnard Jr., Venu Raman and Ishan Barman. Label-free identification of unique metastatic organ-specific human breast cancer signatures featuring Raman spectroscopy. *2nd International Conference on Label-Free Technologies.* March 12-14, 2015, Boston.

* Yanbo Tao, Xinyu Wu, Yong Yang, Huihuan Qian, **Chi Zhang**, and Yangsheng Xu. An Intelligent Shoe System: Evaluation of Weight Load Approaches by Gait Analysis. *2012 International Conference on Computerized Healthcare (ICCH 2012).* December 17-18, 2012, Hong Kong.