HARNESSING LAEBL-FREE RAMAN SPECTROSCOPY

FOR METASTATIC CANCER DIAGNOSIS AND

BIOLOGIC DEVELOPMENT

By
Chi Zhang

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

Baltimore, Maryland
September, 2019

Abstract

Optical spectroscopy is unique amongst experimental techniques in that it can be performed in near-physiological conditions, achieve high molecular specificity, and explore dynamics on timescales ranging from nanoseconds to days. In particular, Raman spectroscopy has emerged in the last two decades as a uniquely versatile method to investigate the structures and properties of molecules in diverse environments through interpreting vibrational transitions.

In this thesis, we present four interconnected biomedical and biopharmaceutical applications of Raman spectroscopy that exploit its exquisite molecular specificity, non-perturbative nature, and near real-time measurement capability.

In the first presented study, we harness spontaneous Raman spectroscopy in conjunction with multivariate analysis to rapidly and quantitatively determine antibody-drug conjugate aggregation with the goal of eventual application as an in-line tool for monitoring protein particle formation. By exploring subtle, but consistent, differences in spectral vibrational modes of various monoclonal antibodies (mAb) aggregations, a support vector machine-based regression model is developed which is able to accurately predict a wide range of protein aggregation. In addition, the investigation of these spectral vibrational modes also offers new insights into mAb product-specific aggregation mechanisms.

Second, leveraging surface-enhanced Raman scattering (SERS) and localized surface plasmon resonance (LSPR), we present a design of plasmonic nanostructures based on rationally structured metal-dielectric combinations, which we call composite scattering probes (CSP). Specifically, we design CSP configurations that have several prominent

resonance peaks enabling higher tunability and sensitivity for self-referenced multiplexed analyte sensing. The CSP prototypes were used to demonstrate differentiation of subtle changes in refractive index (as low as 0.001) as well as acquire complementary untargeted plasmon-enhanced Raman measurements from the biospecimen's compositional contributors.

In the third study, we demonstrate that Raman spectroscopy offers vital biomolecular information for early diagnosis and precise localization of breast cancer-colonized bone alterations. We show that as early as two weeks after intracardiac injections of breast cancer cells in mouse models, Raman measurements in femur and spine uncover consistent changes in both bone matrix and mineral composition. This research effort opens the door for improved understanding of breast metastatic tumor-related bone remodeling and establishing a non-invasive tool for detection of early metastasis and prediction of fracture risk.

In parallel with this effort, we also seek to identify the differences between organ-specific isogenic metastatic breast cancer cells. By interpreting the informative spectral bands, we are able to unambiguously identify these isogenic cell lines as unique biological entities. Our spectroscopic study and corresponding metabolic research indicate that tissue-specific adaptations generate biomolecular alterations on cancer cells.

Advisor: Dr. Ishan Barman
Dissertation Reader: Dr. Tza-Huei (Jeff) Wang and Dr. Yun Chen

Acknowledgments

First, I would like to express my sincere appreciation to my advisor, Dr. Ishan Barman, for accepting me as a Ph.D. student with little background in optics and, subsequently, offering me the opportunity to work on multiple challenging projects. His careful and consistent guidance has always motivated me to face upcoming challenges. In addition, his intense curiosity about new technologies and the willingness to explore uncharted scientific territories will be a beacon for me, now and in the future.

Furthermore, I would like to thank my thesis readers Dr. Jeff Wang and Dr. Yun Chen. I greatly appreciate their valuable suggestions regarding my research work. I would especially like to acknowledge Dr. Chen for letting me use the confocal microscope in her lab for many of my measurements. I also want to take this opportunity to express my gratitude to Dr. Steve Semancik at National Institute of Standards and Technology (NIST). His guidance and support enabled me to fabricate some wonderful nanostructures at the NIST facilities. Special thanks are also due to Dr. Rishikesh Pandey at Connecticut Children's Innovation Center who encouraged me to participate in various research projects and showed me the potential of spectroscopic techniques. I would like also like to express my gratitude to Dr. Venu Raman and Dr. Paul T. Winnard Jr., who generously assisted with the biological interpretation of our spectroscopic data for the cancer metastasis studies. A big thank you to Dr. Jeremy Springall for offering me the wonderful opportunity of interning at MedImmune, LLC (Gaithersburg), an experience that shaped my understanding of the biopharmaceutical industry.

Table of Contents

List of Tables

List of Figures

variations. The shaded regions above the 2DCOS plots highlight (from left to right) the CH deformation, CH_3 band, CH_2 bending and amide-II modes in the Raman spectrum.

background removal. The solid lines depict the mean spectrum with associated shadings representing the ±1 standard deviations (SD). Prominent Raman peaks are indicated in the biologic spectrum. The spectra were normalized to the intensity of the 983 cm^{-1} peak – characteristic of the mobile phase background. The 1160-1800 cm^{-1} (black dashed box) fingerprint region was used for further multivariate model development. Spectral features of CH deformation, CH_2 bending and amide-II modes (from left to right) are marked by yellow and purple dashed lines.

SEM image of the SNPG sample after 20 min of annealing on a hotplate. The inset shows the zoomed-in version marked by the blue dashed box. The gold (yellow) and the silver (blue) regions of the SNPG structure are highlighted in false color. (E) SEM image of the STG sample. The inset shows the zoomed-in version marked by the red dashed box. The gold (yellow) and the silver (blue) regions of the STG structure are indicated by the false color representation. The scale bars for all the SEM images marked in white are 200 nm. (F) s-SNOM image of the annealed SNPG sample; although the silver nanoparticles cannot be resolved, the gold nanoparticles are readily identified (s-SNOM image courtesy: Bruker Nano Surfaces Division).

silver nanostructures are in gray color. (C) Simulated absorption spectra for different numerical configurations are plotted along with absorption spectra of a gold sphere. (D) Experimental absorption spectra of various fabricated samples – 4nm SNPG, 4 nm SNPG after annealing, STG and gold islands. δ signifies the shift in peak position of the SNPG sample due to annealing.

could be suitably classified. (C) Hierarchical cluster analysis of absorption spectra of resonance peak of STG at various concentrations of BSA, air and water. Different solutions could be identified although the accuracy is less compared to that obtained with the SNPG sample. (D) PLS prediction results of BSA solutions, water and air in the wavelength range of 320-500 nm (around the 440 nm peak region) in case of the SNPG-annealed substrate. The solid line denotes y=x and the red dashed box is an enlarged version of a part of the graph as shown in the subfigure. (E) PLS prediction residuals for absorption spectra of SNPG-annealed sample (around the 440 nm peak) belonging to the different RI.

Figure 3.5 **SERS measurements of BSA solution and numerical** 68 **calculation of field enhancement:** (A) SERS spectra of 10 μmol/L BSA on different substrates measured with a 2 second integration with a 530 nm laser in a micro-Raman system. Prominent Raman peaks are marked in the figure. The solid lines depict the average spectrum, and the shaded region represents the $\pm$ standard deviations (SD). (B) Distribution of $|E/E_0|^4$ in the XZ plane at 530 nm excitation. All the panels are plotted with the same range of color scale for comparison. The scale bar is 20 nm. The STG configuration shows significant enhancement at the sharp corners,

which is due to the lightning rod effect but is absent in our experimental prototype.

(left panel) and tumor-bearing mice at 4 and 5 weeks post 435-tdT inoculations. In the week 5 image, fluorescent metastatic lesions in the scapulae (small arrows), lower thoracic-upper lumbar region of the spine (large arrow), and left proximal femur/pelvis region (arrowhead) were evident. Note, to better ascertain bone fluorescence signals, intense brain fluorescence was masked in the middle and right panel images. (B) X-ray images from a Faxitron x-ray scanner displaying the femur of the same mice shown in the corresponding panels in (A). No metastatic bone lesions were revealed in any of the x-ray images.

mineral crystallinity from 1/FWHM PO_4^{3-} (1/FWHM phosphate v1 peak) calculations. Relative to week 0, average compositional changes of (A) femurs and (B) spines at 2, 4, and 5 weeks post tumor cell inoculations were quantified. Error bars = ±1 SD. (n,s, denotes not significant, *$p<0.05$, **$p<0.005$) (C) Distinct clustering of the spectral data corresponding to each week was revealed in the case of the femur analyses while two clusters emerged in the analysis of the spine data, namely, an early stage cluster: week 0 + week 2 and a late stage cluster: week 4 + week 5. Blue circles = week 0, red squares = week 2, green triangles = week 4, and orange diamond = week 5.

week (0, 2, 4, 5) of the study. Autofluorescence was low (week 0 images), metastasis specific fluorescent signals from tdT-435 cells within the metaphysis regions was relatively weak at week 2 and much more intense at week 4 and 5. (B) Fold increases in fluorescent intensities from the metaphysis regions of femurs in (A) relative to week 0 autofluorescence as well as between weeks as determined by the semi-quantitative measurements. Error bars = ±1 SD. Two-tailed Students t-test was employed for evaluating statistical significance (asterisk depicts $p<0.05$).

was apparent at 48 hr post inoculation of the plate with distinct small spherical cells making up each colony (arrow heads) and thin cellular extensions/filopodia (micro- or nanotubes; arrows). (D) After 120 hr the interconnected colony pattern remained. (E–F) Two examples of the characteristic growth pattern at "confluency" of the brain cell line with colonies elaborately linked together by nanotubes. These interconnections between cells/colonies have consistently been recorded at > 100 μm in length. (G) Higher magnification of the central portion of image (E). (H) Expanded image of the lower left-hand corner of image (G). These magnified images allow for a very clear visualization of the complex and intricate web of interconnections between colonies that were in place. Scale bars in all images depict 100 μm.

Expression values for the heat maps are indicated by a key at the bottom of the maps.

Figure 5.6 **Raman spectroscopic analyses of organ-specific metastatic breast cancer cell lines reveals distinct spectral characteristics for each cell line.** (A) Representative Raman spectra acquired from brain, primary tumor (1° Tumor), liver, lung, and spine cell lines. The solid profile depicts the mean spectrum of each sample group and the shadow represents ±1 standard deviation. Spectra were normalized and offset for visualization. Dashed vertical lines delineate Raman shifts (cm^{-1}) detailed in **Table 5.3**. (B) Principal component (PC) loadings for PC 1, 2, 3 and 5, for the Raman measurements are shown. Dashed vertical lines delineate prominent Raman shifts (cm^{-1}) detailed in **Table 5.3**. (C) Radial visualization principal component scores plot, corresponding to the most discriminative PCs (PC 1, 2, 3, and 5), shows the clustering of the spectral data corresponding to each organ-specific cell line, red: primary tumor, blue: brain, green: liver, orange: lung, and purple: spine. (D) Dendrogram of organ-specific breast cancer cell lines cluster analysis. Each color bar represents one organ-specific cell line. (E) Identification of informative spectral regions via PCA data exploration as exemplified by the PC loadings corresponding to the spectral dataset acquired from: primary tumor and liver (left panel) and primary tumor and spine (right panel) cell lines. The top

to bottom profiles in each panel show difference spectra: (DS) between liver/primary or spine/primary spectra along with their PC 1 and PC 2 loadings, respectively. The highlighted yellow bars (1–4), represent the wavelength regions elucidated from the difference spectra (DS) as those with the most significant variability amongst the considered cell lines.

phenylethylamine. Profiles in blue represent the difference spectra: (DS) whereas the red and green profiles show the PC 1 and PC 2 loadings respectively for each chosen pair of cell lines.

Chapter 1. Introduction

1.1 What is Raman spectroscopy

Raman spectroscopy is a fundamental form of molecular spectroscopy that exploits the inelastic scattering of light, which was first experimentally observed by C.V. Raman in 1928s[1]. Briefly, when a photon interacts with a molecule, the molecule can be excited into a transient virtual energy state. The electrons at this virtual energy state can only exist for a very short time, usually 10^{-12} second[2], before they relax back to the initial ground state by releasing photons with the same energy as the excitation. This process is known as elastic (Rayleigh) scattering. However, there is a small but finite possibility, *ca.* 1 in 10^8 interacted photons, that the light-matter interaction occurs in an inelastic manner by scattering photons with different energy compared with the incident photons. This phenomenon is known as Raman scattering[3]. Under different nuclear motions, the energy difference between the incident photon and released photon could be positive or negative. If the energy of released photons is lower than the incident photons, the scattering is known as Stokes Raman scattering. In contrast, for anti-Stokes Raman scattering, the scattered photons have greater energy than incident photons[4,5].

The energy differences between the incident light and Raman scattering light, which is typically presented in intensity versus wavenumber plots, directly reflect the specific molecular bond vibration status. In addition, the intensity of the Raman spectral feature(s) is linearly related to the molecule's concentration. These attributes make Raman spectroscopy a powerful label-free analytical tool for fingerprinting and quantitation of molecules. Furthermore, Raman measurements can be made in near real-time, and do not

necessitate substantial sample preparation thereby ensuring minimum disturbance to the native condition of the interrogated sample[3].

Although Raman scattering was first observed nearly a century ago, its application particularly in biomedical research and pharmaceutical development has attracted significant attention, due to the spectacular advances in laser and charge-coupled device technology. A number of related analytical tools, including but not limited to resonance Raman spectroscopy (RRS), surface-enhanced Raman spectroscopy (SERS), spatially offset Raman spectroscopy (SORS), coherent anti-stokes Raman spectroscopy (CARS) and stimulated Raman scattering (SRS), have also substantial development.

Spontaneous Raman spectroscopy is the first and, perhaps, still the most commonly employed Raman scattering technique used in biomedical analyses. Combined with light microscopy, spontaneous Raman spectroscopy allows the generation of high-resolution composite images that bridges the biomolecular and morphological domains. Several publications have been reported with this such spectroscopic imaging to visualize the distribution of chemical components, and to identify the biochemical alteration regions, for instance in tumor-bearing tissue[6-8].

The low likelihood of Raman scattering determines the intrinsic weak signal intensity. Researchers have worked hard to overcome this disadvantage and one solution is using RRS. Theoretically, Raman scattering could happen using any excitation wavelength. However, selection of a excitation line corresponding to the electronic absorption maximum of the target analytes substantially enhances the intensity of those specific Raman bands (amplification of $\sim 10^3 - 10^5$)[9-11]. RRS has been applied to study mitochondria

cytochrome-c release during cell apoptosis[12,13], and to investigate the solvent-dependent coexistence of localized & delocalized dinitroaromatic radical anions[14].

Besides RRS, SERS provides another solution to enhance the weak spontaneous Raman signal. Instead of changing the hardware of the spectroscopic system, SERS employs noble metal nanoparticles or substrates with the rough metal surface to generate localized surface plasmons. Depending on the type of nanoparticle or nano-structured substrate, the signals of adsorbed molecules (or those in the vicinity) can be enhanced by a factor of 10^6 to 10^{10} [5,15-17]. The distance between the metal surface and analyte is the most dominant factor of signal enhancement, and the enhancement factor decays exponentially with distance increase[18-21]. It is worth noting that the fluorescence signal coming from the analyte(s), which often dominates the weak Raman signatures, gets quenched when the molecules are close to the noble metal surface while the Raman signal is significantly enhanced[22-25]. In terms of noble metal selection for nanoparticle synthesis or nanostructure fabrication in the biological field, gold is the first choice due to its good biocompatibility[26,27]. In addition, gold nanoparticles can be readily functionalized by DNA and antibodies through thiol group binding[28-30]. Recently, functionalized gold nanoparticles-Raman reporter combination has been applied in cancer diagnosis, bioimaging, and photothermal therapy[31-33].

For tissue analyses, a significant challenge for optical measurements (Raman scattering or otherwise) is overcoming the limited light penetration depth. It is challenging to acquire biochemical information from deeper layers of the tissue. One possible solution is using near-infrared laser, as there is less absorption of light by water in the tissue at this wavelength range[34,35]. Another recently proposed solution for this

3

problem is utilizing SORS that advocates the collection of diffuse Raman scattering photons at different spatial offsets from the laser excitation spot. A SORS measurement consist of at least two Raman measurements; one at the source and one at an offset position of typically a few millimeters away. The two acquired spectra can be subjected using a scaled subtraction to generate two spectra representing the subsurface and surface spectra. With help from SORS, it is possible to acquire biochemical information from 40 mm below the sample surface[36,37], which is much deeper than that possible using near-infrared light alone (approximate 5 mm)[38,39]. Due to this deeper imaging ability, SORS has been employed in diverse applications such as recovering hidden images in art as well as in non-invasive bone disease and breast cancer diagnosis[40,41]. In addition, SORS can be coupled with SERS to retrieve Raman signals from even deeper layers in diffusely scattering samples. This joint technique first reported by Stone and co-workers is called surface-enhanced SORS (SESORS)[42,43]. Similar to SORS, SESORS has a wide range of application in non-invasive detection field, such as *in vivo* glucose sensing and brain tissue characterization through the skull in a non-invasive manner[44,45].

Unlike other Raman-related techniques that use a single wavelength laser for excitation, CARS and SRS require two pulsed lasers with different wavelengths to reveal the chemical information from samples with the help of nonlinear optical processes[46,47]. By tuning the wavelength differences between the pump laser and Stokes laser to match a particular sample's vibrational mode, the Raman signal intensity is significantly increased. Unlike spontaneous Raman and RRS, CARS and SRS have a clean spectral background without fluorescence interference. Furthermore, the nonlinear nature of

CARS and SRS opens the door for three-dimensional sectioning and for deeper tissue imaging[48,49].
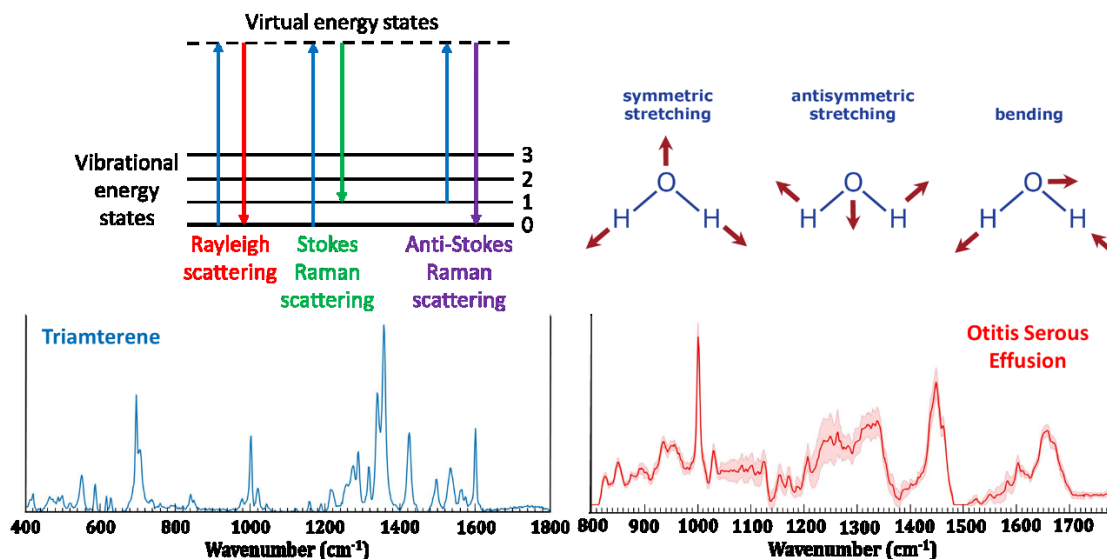
Figure 1.1 Schematic of Raman scattering, molecular band vibration and representative Raman spectra acquired from triamterene (a diuretic) and serous otitis media.

1.2 Specific research questions and thesis outline

This dissertation is organized in the following manner, as illustrated in **Figure 1.2**.

In Chapter 2, we discuss our efforts to build a rapid, quantitative tool for determination of aggregation and particle formation in antibody drug conjugate therapeutics. The contents of this work have been recently published in *Analytica Chimica Acta* (2019). The recent shift of therapeutic focus towards antibody-based drugs has led to continuous discovery and development of new biologics, and concerted efforts in their rapid translation to the clinic. While the commercialization of monoclonal antibodies (mAbs) is gathering momentum, addressing critical challenges in manufacturing and quality control of these products is vital to the long-term success of this sector. One of the principal quality control challenges is stability testing of mAbs, particularly protein aggregation characterization, in

order to ensure therapeutic material dosed to patients is safe and efficacious, and consistent with previous clinical and toxicological experience. While important, real-time identification and quantification of subvisible particles in the mAb drug products remain inaccessible with existing techniques due to limitations in measurement time, sensitivity or experimental conditions. We propose to harness Raman spectral markers to offer a new route for real-time, quantitative determination of aggregation and to understand how physiochemical properties may impact the formation of high molecular weight species. Meeting these needs also requires the ability to deal with complex spectral datasets that encode underlying multivariate interactions. Here, we report the application of spontaneous Raman spectroscopy coupled with multivariate data analysis for predicting aggregation level of mAbs – with minimal sample preparation and in a near real-time manner.

In Chapter 3, we report composite-scattering plasmonic nanoprobes for label-free, quantitative biomolecular detection that combine SERS with localized surface plasmon resonance (LSPR) sensing. This chapter is reproduced with permission from *Small* (2019), where the findings of this study were published. Despite the significant advances in the development of SERS nanostructured substrate, the potential of core-shell nanoparticle-based designs for enabling both SERS and LSPR sensing has been has been surprisingly underappreciated - even though such constructs have received attention as a way to tune plasmon resonance and electromagnetic enhancement. Using simple microfabrication techniques, notably physical vapor deposition and thermal de-wetting, we have successfully fabricated a uniform, large-area plasmonic substrate with facile spectroscopic tunability. Combining spectroscopic data with partial least squares has enabled resolving refractive index differences as low as 0.001. The refractive index related change in the

absorption spectra and SERS measurements, using albumin as a model analyte, closely agree with our finite element method-based simulations for different configuration of the nanostructures.

Chapter 4 focuses on the development of Raman spectroscopy as a label-free, non-invasive tool for recognizing breast cancer metastases to the bone significantly before it is detected by existing imaging technologies. This research effort was published in *Chemical Science* (2018) and is reproduced here from the published article. Our research aims to realize automated and objective cancer diagnoses and grading that betters human performance, and to establish early cancer detection not detectable by current pathologic examinations. Epithelial cancers including breast and prostate commonly progress to form incurable bone metastases. For this to occur, breast cancer cells must adapt their phenotype and behavior to enable detachment from the primary tumor, invasion into the vasculature, and homing to and subsequent colonization of bone. Yet, existing imaging tools offer frustratingly little information on the bone's biochemical and mineral composition and whether it has been compromised. Here, by studying femurs and spines of mouse models inoculated with fluorescent-expressing MDA-MB-435 cells that recapitulate spontaneous breast cancer dissemination to the bone, we reveal that Raman spectroscopy has the capability to detect biochemical changes to the structure of bones associated with early cancer metastasis without *a priori* imaging or pathological knowledge of lesion location. Our findings create a new landscape for Raman spectroscopic monitoring by: (A) uncovering discriminative spectroscopic features that map to molecular alterations in the bone following breast cancer metastasis; (B) offering quantitative measures to determine longitudinal and location-specific osteolytic/osteoblastic variations caused by the tumor

involvement; and (C) demonstrating the feasibility of using Raman spectroscopy for early detection of metastatic disease in bone.

In Chapter 5, we present, to the best of our knowledge, the first demonstration that a label-free spectroscopic technique can be used to assess organ-specificity of distant metastases. Identifying distinct molecular signatures of matched primary breast cancer and metastatic lesions with the principal goal of designing therapeutic regimens directed at eradicating lethal metastatic disease is an ongoing investigation. Currently the preclinical validation of this process has been based on generating metastatic lesions using either tail-vein or intra-cardiac injections of breast cancer cells. Subsequent genomic and proteomic analyses of the isolated metastatic cell lines have been relied upon as guidelines to profile potential therapeutic options. However, this approach has generated inconsistent data sets that have undermined the development of successful clinical treatments. Our findings, originally published in *Oncotarget* (2017), underscore the importance of establishing an appropriate preclinical breast cancer metastasis model that takes into account the effect of organ-specific microenvironments on the differential evolution of tumor cells.

In Chapter 6, the major findings and conclusions of our work are summarized and the future directions outlined. We believe that further research along these specific directions, coupled with the work presented in this thesis, will enable further development and translation of this powerful spectroscopic tool to the clinic.
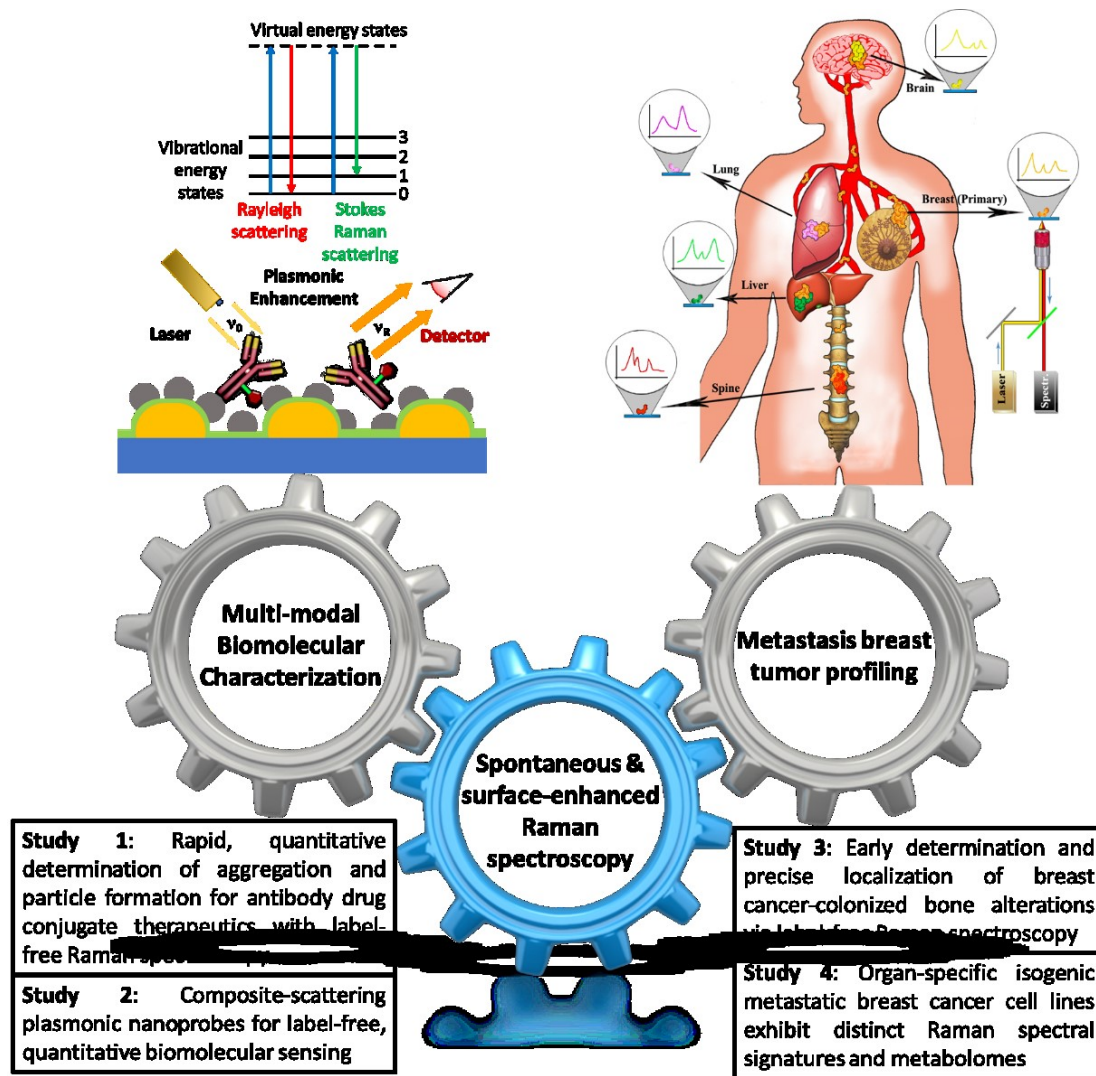
Figure 1.2 **Schematic representation of the structure of this thesis.** The reported Raman spectroscopic applications could be classified as molecular fingerprints characterization and biological tissue profiling measurement. The spontaneous Raman spectroscopy and surface-enhanced Raman spectroscopy are discussed in corresponding chapters.

Bibliography

1. Raman, C.V. and Krishnan, K.S. A new type of secondary radiation. *Nature*, **1928,** 121, pp.501.

2. Kang, K., Abdula, D., Cahill, D.G. and Shim, M. Lifetimes of optical phonons in graphene and graphite by time-resolved incoherent anti-Stokes Raman scattering. *Physical Review B*, **2010**, 81, pp.165405.

3. Ember, K.J., Hoeve, M.A., McAughtrie, S.L., Bergholt, M.S., Dwyer, B.J., Stevens, M.M., Faulds, K., Forbes, S.J. and Campbell, C.J. Raman spectroscopy and regenerative medicine: a review. *NPJ Regenerative medicine*, **2017**, 2, pp.12.

4. Smith, E. and Dent, G. *Modern Raman spectroscopy: a practical approach*. Wiley, **2019**.

5. Le Ru, E. and Etchegoin, P. *Principles of Surface-Enhanced Raman Spectroscopy: and related plasmonic effects*. Elsevier, **2008**.

6. Krafft, C., Knetschke, T., Siegner, A., Funk, R.H. and Salzer, R. Mapping of single cells by near infrared Raman microspectroscopy. *Vibrational Spectroscopy*, **2003**, 32, pp.75-83.

7. Krishna, C.M., Sockalingum, G.D., Kegelaer, G., Rubin, S., Kartha, V.B. and Manfait, M. Micro-Raman spectroscopy of mixed cancer cell populations. *Vibrational Spectroscopy*, **2005**, 38, pp.95-100.

8. Chan, J.W., Lieu, D.K., Huser, T. and Li, R.A. Label-free separation of human embryonic stem cells and their cardiac derivatives using Raman spectroscopy. *Analytical chemistry*, **2009**, 81, pp.1324-1331.

9. Smith, E. and Dent, G. *Modern Raman spectroscopy: a practical approach*. Wiley, **2019**.

10. Ehlerding, A., Johansson, I., Wallin, S. and Östmark, H. Resonance-enhanced Raman spectroscopy on explosives vapor at standoff distances. *International Journal of Spectroscopy*, **2012**.

11. Zardi, P., Gallo, E., Solan, G.A. and Hudson, A.J. Resonance Raman spectroscopy as an in situ probe for monitoring catalytic events in a Ru–porphyrin mediated amination reaction. *Analyst*, **2016**, 141, pp.3050-3058.

12. Okada, M., Smith, N.I., Palonpon, A.F., Endo, H., Kawata, S., Sodeoka, M. and Fujita, K. Label-free Raman observation of cytochrome c dynamics during apoptosis. *Proceedings of the National Academy of Sciences*, **2012**, 109, pp.28-32.

13. Kitt, J.P., Bryce, D.A., Minteer, S.D. and Harris, J.M. Raman spectroscopy reveals selective interactions of cytochrome c with cardiolipin that correlate with membrane permeability. *Journal of the American Chemical Society*, **2017**, 139, pp.3851-3860.

14. Hoekstra, R.M., Chen, Y.T., Kiesz, M.D., Telo, J.P., Stephenson, R.M., Nelsen, S.F. and Zink, J.I. Resonance Raman spectroscopic study of solvent-dependent coexistence of localized and delocalized dinitroaromatic radical anions. *Canadian Journal of Chemistry*, **2014**, 92, pp.940-947.

15. Kneipp, K., Wang, Y., Kneipp, H., Perelman, L.T., Itzkan, I., Dasari, R.R. and Feld, M.S. Single molecule detection using surface-enhanced Raman scattering (SERS). *Physical review letters*, **1997**, 78, p.1667.

16. Le Ru, E.C., Blackie, E., Meyer, M. and Etchegoin, P.G. Surface enhanced Raman scattering enhancement factors: a comprehensive study. *The Journal of Physical Chemistry C*, **2007**, 111, pp.13794-13803.

17. Stiles, P.L., Dieringer, J.A., Shah, N.C. and Van Duyne, R.P. Surface-enhanced Raman spectroscopy. *Annu. Rev. Anal. Chem.*, **2008**, 1, pp.601-626.

18. Etchegoin, P.G. and Le Ru, E.C. A perspective on single molecule SERS: current status and future challenges. *Physical Chemistry Chemical Physics*, **2008**, 10, pp.6079-6089.

19. Fleischmann, M., Hendra, P.J. and McQuillan, A.J. Raman spectra of pyridine adsorbed at a silver electrode. *Chemical physics letters*, **1974**, 26, pp.163-166.

20. Albrecht, M.G. and Creighton, J.A. Anomalously intense Raman spectra of pyridine at a silver electrode. *Journal of the american chemical society*, **1977**, 99, pp.5215-5217.

21. Jeanmaire, D.L. and Van Duyne, R.P. Surface Raman spectroelectrochemistry: Part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode. *Journal of electroanalytical chemistry and interfacial electrochemistry*, **1977**, 84, pp.1-20.

22. Li, M., Gou, H., Al-Ogaidi, I. and Wu, N. Nanostructured sensors for detection of heavy metals: a review. *Sustainable Chem. Eng.* **2013**, 1, pp.713-723.

23. Shan, F., Zhang, X.Y., Fu, X.C., Zhang, L.J., Su, D., Wang, S.J., Wu, J.Y. and Zhang, T. Investigation of simultaneously existed Raman scattering enhancement and inhibiting fluorescence using surface modified gold nanostars as SERS probes. *Scientific reports*, **2017**, 7, p.6813.

24. Walters, C.M., Pao, C., Gagnon, B.P., Zamecnik, C.R. and Walker, G.C. Bright Surface-Enhanced Raman Scattering with Fluorescence Quenching from Silica Encapsulated J-Aggregate Coated Gold Nanoparticles. *Advanced Materials*, **2018**, 30, p.1705381.

25. Han, X.X., Zhao, B. and Ozaki, Y. Surface-enhanced Raman scattering for protein detection. *Analytical and bioanalytical chemistry*, **2009**, 394, pp.1719-1727.

26. Daniel, M.C. and Astruc, D. Gold nanoparticles: assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chemical reviews*, **2004**, 104, pp.293-346.

27. Qian, X., Peng, X.H., Ansari, D.O., Yin-Goen, Q., Chen, G.Z., Shin, D.M., Yang, L., Young, A.N., Wang, M.D. and Nie, S. In vivo tumor targeting and spectroscopic detection with surface-enhanced Raman nanoparticle tags. *Nature biotechnology*, **2008**, 26, pp.83.

28. Jazayeri, M.H., Amani, H., Pourfatollah, A.A., Pazoki-Toroudi, H. and Sedighimoghaddam, B. Various methods of gold nanoparticles (GNPs) conjugation to antibodies. *Sensing and bio-sensing research*, **2016**, 9, pp.17-22.

29. Hinman, S.S., McKeating, K.S. and Cheng, Q. DNA linkers and diluents for ultrastable gold nanoparticle bioconjugates in multiplexed assay development. *Analytical chemistry*, **2017**, 89(7), pp.4272-4279.

30. Kim, E.Y., Stanton, J., Vega, R.A., Kunstman, K.J., Mirkin, C.A. and Wolinsky, S.M. A real-time PCR-based method for determining the surface coverage of thiol-capped oligonucleotides bound onto gold nanoparticles. *Nucleic acids research*, **2006**, 34, pp.e54.

31. Wang, Z., Zong, S., Wang, Y., Li, N., Li, L., Lu, J., Wang, Z., Chen, B. and Cui, Y. Screening and multiple detection of cancer exosomes using an SERS-based method. *Nanoscale*, **2018**, 10, pp.9053-9062.

32. Jin, X., Khlebtsov, B.N., Khanadeev, V.A., Khlebtsov, N.G. and Ye, J. Rational design of ultrabright SERS probes with embedded reporters for bioimaging and photothermal therapy. *ACS applied materials & interfaces*, **2017**, 9, pp.30387-30397.

33. Zhao, J., Zhang, K., Li, Y., Ji, J. and Liu, B. High-resolution and universal visualization of latent fingerprints based on aptamer-functionalized core–shell nanoparticles with embedded SERS reporters. *ACS applied materials & interfaces*, **2016**, 8, pp.14389-14395.

34. Tsai, C.L., Chen, J.C. and Wang, W.J. Near-infrared absorption property of biological soft tissue constituents. *Journal of Medical and Biological Engineering*, **2001**, 21, pp.7-14.

35. Shi, L., Sordillo, L.A., Rodríguez-Contreras, A. and Alfano, R. Transmission in near-infrared optical windows for deep brain imaging. *Journal of biophotonics*, **2016**, 9, pp.38-43.

36. Ghita, A., Matousek, P. and Stone, N. Exploring the effect of laser excitation wavelength on signal recovery with deep tissue transmission Raman spectroscopy. *Analyst*, **2016**, 141, pp.5738-5746.

37. Matousek, P. and Stone, N. Development of deep subsurface Raman spectroscopy for medical diagnosis and disease monitoring. *Chemical Society Reviews*, **2016**, 45, pp.1794-1802.

38. Stolik, S., Delgado, J.A., Perez, A. and Anasagasti, L. Measurement of the penetration depths of red and near infrared light in human "*ex vivo*" tissues. *Journal of Photochemistry and Photobiology B: Biology*, **2000**, 57, pp.90-93.

39. Ash, C., Dubec, M., Donne, K. and Bashford, T. Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers in medical science*, **2017**, 32, pp.1909-1918.

40. Botteon, A., Conti, C., Realini, M., Colombo, C. and Matousek, P. Discovering hidden painted images: subsurface imaging using microscale spatially offset Raman spectroscopy. *Analytical chemistry*, **2016**, 89, pp.792-798.

41. Matousek, P. Spatially offset Raman spectroscopy for non-invasive analysis of turbid samples. *TrAC Trends in Analytical Chemistry*, **2018**, 103, pp.209-214.

42. Stone, N., Faulds, K., Graham, D. and Matousek, P. Prospects of deep Raman spectroscopy for noninvasive detection of conjugated surface enhanced resonance Raman scattering nanoparticles buried within 25 mm of mammalian tissue. *Analytical chemistry*, **2010**, 82, pp.3969-3973.

43. Stone, N., Kerssens, M., Lloyd, G.R., Faulds, K., Graham, D. and Matousek, P. Surface enhanced spatially offset Raman spectroscopic (SESORS) imaging–the next dimension. *Chemical Science*, **2011**, 2, pp.776-780.

44. Ma, K., Yuen, J.M., Shah, N.C., Walsh Jr, J.T., Glucksberg, M.R. and Van Duyne, R.P. In vivo, transcutaneous glucose sensing using surface-enhanced spatially offset Raman spectroscopy: multiple rats, improved hypoglycemic accuracy, low incident power, and continuous monitoring for greater than 17 days. *Analytical chemistry*, **2011**, 83, pp.9146-9152.

45. Sharma, B., Ma, K., Glucksberg, M.R. and Van Duyne, R.P. Seeing through bone with surface-enhanced spatially offset Raman spectroscopy. *Journal of the American Chemical Society*, **2013**, 135, pp.17290-17293.

46. Evans, C.L. and Xie, X.S. Coherent anti-Stokes Raman scattering microscopy: chemical imaging for biology and medicine. *Annu. Rev. Anal. Chem.*, **2008**, 1, pp.883-909.

47. Camp Jr, C.H. and Cicerone, M.T. Chemically sensitive bioimaging with coherent Raman scattering. *Nature photonics*, **2015**, 9, p.295.

48. Zumbusch, A., Holtom, G.R. and Xie, X.S. Three-dimensional vibrational imaging by coherent anti-Stokes Raman scattering. *Physical review letters*, **1999**, 82, p.4142.

49. Yakovlev, V.V., Zhang, H.F., Noojin, G.D., Denton, M.L., Thomas, R.J. and Scully, M.O. Stimulated Raman photoacoustic imaging. *Proceedings of the National Academy of Sciences*, **2010**, 107, pp.20335-20339.

Chapter 2. Rapid, quantitative determination of aggregation and particle formation for antibody drug conjugate therapeutics with label-free Raman spectroscopy

2.1 Abstract

Lot release and stability testing of biologics are essential parts of the quality control strategy for ensuring therapeutic material dosed to patients is safe and efficacious, and consistent with previous clinical and toxicological experience. Characterization of protein aggregation is of particular significance, as aggregates may lose the intrinsic pharmaceutical properties as well as engage with the immune system instigating undesirable downstream immunogenicity. While important, real-time identification and quantification of subvisible particles in the monoclonal antibody (mAb) drug products remains inaccessible with existing techniques due to limitations in measurement time, sensitivity or experimental conditions. Here, owing to its exquisite molecular specificity, non-perturbative nature and lack of sample preparation requirements, we propose label-free Raman spectroscopy in conjunction with multivariate analysis as a solution to this unmet need. By leveraging subtle, but consistent, differences in vibrational modes of the biologics, we have developed a support vector machine-based regression model that provides fast, accurate prediction for a wide range of protein aggregations. Moreover, in blinded experiments, the model shows the ability to precisely differentiate between aggregation levels in mAb like product samples pre- and post-isothermal incubation, where

an increase in aggregate levels was experimentally determined. In addition to offering fresh insights into mAb like product-specific aggregation mechanisms that can improve engineering of new protein therapeutics, our results highlight the potential of Raman spectroscopy as an in-line analytical tool for monitoring protein particle formation.

2.2 Introduction

Since their first licensing for clinical use nearly three decades ago, monoclonal antibodies (mAbs) have offered a powerful therapeutic route for targeting specific mutations and defects in protein structure and expression. The high specificity and affinity of mAbs have catalyzed their development for treating a wide range of pathologies[1], such as cancers, infectious diseases and inflammatory conditions, making them the fastest growing group of biotechnology-derived molecules in clinical trials[2]. By the end of 2017, 57 mAbs and 11 biosimilars had been approved by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA)[3] with the global value of the market estimated to be $20 billion per year[4].

However, the production of these therapeutic antibodies requires the use of very large cultures of mammalian cells followed by extensive purification steps leading to extremely high production costs. This is exacerbated by the lack of suitable metrology tools for rapid characterization of key attributes of the biologic product that are directly linked to its safety and efficacy. For instance, there is a critical unmet need to rapidly assess antibody stability during the development and manufacturing phases of a mAb product. Of all possible instabilities, protein aggregation presents a singular challenge. In addition to the mAb-specific aggregation propensity[5], the interplay of physicochemical parameters such as

protein and ion concentrations, particulate contamination, pH, and temperature plays a key role in inducing aggregation in therapeutic formulations. Severe protein aggregation (resulting in protein particles) could lead mAbs to lose their pharmaceutical properties, hinder various upstream/downstream processes, and even stimulate immune response in patients causing harmful effects[6]. Although the latter response is poorly defined, there is emerging evidence of a difference in immune response for aggregated material in comparison to non-aggregated material[7,8]. By achieving rapid and accurate aggregation evaluation, the downstream purification of biopharmaceuticals can be optimized in real-time with the ultimate goal of enhancing the product quality during manufacturing campaigns.

High-performance size exclusion chromatography (HP-SEC) is widely employed for detailed characterization of therapeutic proteins and is often considered as the reference method for qualitative and quantitative evaluation of aggregates[9-11]. Besides HP-SEC, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)[12], asymmetrical field flow fractionation (AF4)[13], fluorescence spectroscopy[14], circular dichroism (CD)[15,16] and dynamic light scattering (DLS)[17] are also applied for quantification of protein aggregation. In addition, mass spectrometry and its variants are extensively used to confirm the fragment masses and, increasingly, to characterize oligomeric protein aggregates[18,19]. Although these techniques have been employed for protein aggregation characterization under various conditions, each has well-known limitations. For instance, mass spectrometry can be time-consuming and requires substantial expertise to properly execute. Meanwhile, DLS has relatively low sensitivity with the obtained size distribution being biased towards larger particles owing to the dependence of the intensity on the sixth power

of the diameter[20]. This is particularly limiting in quantifying solutions with low aggregation content. On the other hand, additional (often complicated) sample-specific preparation is needed for enabling fluorescence measurements. For the most widely employed technique, HP-SEC, the major limitation is that the run times can be long making the analysis time-consuming and difficult to employ as an in-line mAb evaluation tool. Overall, current limitations of the analytical tools used to detect subvisible particles in the mAb drug products make meaningful measurements challenging with different methods often offering widely conflicting results[21].

Development of an in-line method to support real-time analysis requires a rapid, label-free technique, which can detect and quantify low aggregation levels under standard mAb manufacturing conditions. We propose label-free Raman spectroscopy (RS) as a solution to this unmet need based on the wealth of molecular information encoded in the vibrational spectrum. Protein secondary and tertiary structural information are extensively studied through analysis of Raman peak positions and ratios of spectral features that characterize amide-I, amide-II, amide-III, and other backbone vibrations[22-24]. It has also been successfully leveraged to investigate conformational changes in single protein crystals[25] as well as to monitor lyophilization[26] and side chain confirmation[27]. While RS has been previously reported for the elucidation of conformational transitions and aggregation mechanisms in antibodies[28-31], its quantitative power has yet to be utilized for real-time determination of aggregation and to reveal how physiochemical properties may impact the formation of high molecular weight species.

Here, we seek to significantly extend the RS approach by combining it with multivariate analysis of frequent patterns expressed in the spectral profiles for predicting aggregation

level of mAb like products – with minimal sample preparation and in a near real-time manner. For this pilot study, antibody drug conjugate (ADC) samples were first measured by HP-SEC to determine the initial aggregation level. ADCs, which are mAbs attached to cytotoxic drugs by chemical linkers with labile bonds, present a particularly important and understudied cohort for determining protein aggregation. Following fractionation of the initial ADC sample solution a series of mixture samples was established with various levels of aggregate content. Raman spectral data were acquired from these characterized mixture samples using a spectrometer customized for solution measurements. In addition, part of the initial ADC sample solution was stressed at 40°C for a month to rapidly generate product degradation at a faster rate than the anticipated storage condition. Spectroscopic measurements were subsequently performed on the paired set of pre- and post-stress samples.

Two separate studies were performed with the acquired data: (A) Mixture sample study: this featured cross-validation and blinded prediction on a series of ADC mixture samples with different aggregation levels; and (B) Thermal stressing study involving independent prediction of protein aggregation level on ADC samples pre- and post-isothermal incubation. For study (A), principal component analysis (PCA) and 2D correlation analysis (2DCOS) was first used to reveal the subtle, but consistent, spectral changes observed with increasing proportion of mAb like product aggregation. Without any prior knowledge of the amino acid sequence of the tested antibodies, we next developed a support vector machine (SVM)-based regression algorithm for estimation of a wide range of protein aggregation after training it with established aggregation mixture samples. For the thermal stressing study (B), the SVM-derived decision algorithm developed in former study

was imported to predict the (differential) levels of aggregation in the pre- and post-isothermal incubation ADC samples. The collective findings of our study pave the way for real-time, in-line mAb aggregation measurements with RS and, when viewed together with recent reports[31-34], support the expanded use of vibrational spectroscopy as a process analytical tool in the development and manufacturing of protein drugs and biosimilars.

2.3 Experimental section

Sample Preparation. To achieve the goals outlined for studies (A) and (B), multiple mAb samples were derived from an antibody-drug conjugate (ADC) that was kindly provided by MedImmune, LLC (Gaithersburg, MD, USA). Since the aggregation mechanism (and the corresponding spectral changes) are highly mAb-specific (as demonstrated in earlier reports[31]), we have focused on developing an accurate classifier and deriving mechanistic insights for aggregation determination for this biologic sample set. The parent ADC sample was first tested by high-performance size exclusive chromatography (HP-SEC) to determine the protein aggregation level. To facilitate the thermal stressing study, 10 mL of the original sample was incubated at 40°C for one month. The 40°C/75% relative humidity 1-month treatment is a recognized accelerated stability condition from the ICH Q5C[35]. In contrast to some of the prior studies that involved Raman measurements of protein aggregation[28-31], the temperature used in this investigation is significantly below the melting temperature (T_m) of typical mAbs and, hence, is not expected to produce any alteration of the secondary-tertiary structure of the protein molecules[36]. Using HP-SEC, protein monomers and high molecular weight species (HMWS) fractions were collected separately via peak fractionation. Various samples with

0%, 2%, 5%, 10%, 15%, 20%, 30%, 40% and 50% of HMWS were then generated by mixing purified monomer and HMWS samples. Samples were stored at 4°C until spectroscopic measurements. The sample solution was held in 100 μL fused quartz cuvettes and 10 mg/mL concentration was used for all samples. For the protein aggregation mixture samples, the dilution buffer was the HP-SEC mobile phase. On the other hand, the dilution buffer for the original unstressed and 40°C stressed sample was phosphate-buffered saline (PBS). Sample preparation and Raman spectral acquisition were repeated thrice to assess reproducibility of the measurements.

Instrumentation. An Agilent 1260 Infinity HP-SEC was used to characterize the samples' aggregation level. The column used for HP-SEC was TSKgel® G3000SW$_{XL}$ (L × I.D. 30 cm × 7.8 mm, 5 μm particle size) (King of Prussia, PA, USA) and each injection contained 250 μg protein sample with 1 mL/min mobile phase flow. Before each SEC experiment, PBS and gel filtration standard (GFS) were measured to verify system condition.

Raman spectra were acquired by using a μ-ChiralRAMAN-2X Raman spectrometer (ChiralRAMAN, Jupiter, FL, USA). Sample excitation was achieved by using a 532 nm diode laser (MPC6000, Laser Quantum, Fremont, CA, USA). The laser beam passed through a polarizer, a degree of circularity converter and two synchronized counter-rotating half-wave plates. The backscattered Raman signals were transmitted through a notch filter to remove the Raleigh scattered photons, and the spectra were collected using a thermoelectrically cooled CCD camera (MityCCD E3011Bl-DVM, Critical Link, Syracuse, NY, USA). Laser intensity at the sample was kept constant at 50 mW for all the

measurement and the spectral resolution of the system was 7 cm^{-1}. The exposure time for each spectral acquisition was 4 seconds.

Data Analysis. The collected Raman spectra were imported into the MATLAB 2017a (Mathworks, Inc., Natick, MA, USA) environment for further analysis. The Raman wavenumber calibration was performed by using the neon lap spectra acquired by the same system. Spectra were processed to remove interference from cosmic rays. The spectra were subjected to a fifth order best-fit polynomial-based fluorescence removal.

Support vector machines (SVM) were used to develop a regression algorithm for spectroscopically predicting the aggregation level of the proteins. SVM is a supervised learning model that is built on structural risk minimization concepts and can efficiently perform non-linear regression by implicitly mapping the inputs into high-dimensional feature spaces through a kernel. A radial basis function (RBF) with a Gaussian envelope was chosen as the kernel, and the kernel parameters were optimized based on an automated grid search algorithm[37-39].

To examine the predictive power of the SVM-derived regression algorithm, leave-many-out cross-validation and blinded sample tests were performed on the mixture samples in study (A). For the cross-validation tests on the mixture samples, spectra were normalized to the intensity of the 983 cm^{-1} peak, that is characteristic of the HP-SEC mobile phase background. Subsequently, 50% of the spectra from each type of mixture sample were randomly chosen to build the training set. The rest of the spectra were used for cross-validation of the regression model. The algorithm was iterated 100 times with various division of spectra into training and test set to avoid any potential bias. Moreover, for the blinded predictions, the validation set consisted of separate samples with aggregation levels

unknown to the developed regression model. In study (B), the mAb samples were subjected to thermal treatment. These samples did not contain the mobile phase component and, hence, the acquired spectra were normalized to 1643 cm^{-1} amide-I peak. Here, the algorithm developed in (A) was directly used for predicting the HMWS content in the pre- and post-thermal incubation samples without any alteration.

To ensure transferability of algorithms between the two studies, the 1160-1800 cm^{-1} fingerprint region, which is free from the HP-SEC mobile phase interference, was used for multivariate model development. The aggregation levels predicted from SVM-derived regression algorithms were validated against the results from HP-SEC measurements.

Additionally, to better visualize potential correlations between the HMWS content and spectral changes, PCA was employed. PCA is a widely used dimensional reduction technique, which aims to capture the variance in the spectral dataset using only a few orthogonal components (known as principal components, PCs)[40]. Radviz and VizRank algorithms from Orange data mining software[41] were used together with the PC scores to elucidate trends within the spectral data.

Furthermore, in order to gain a better understanding of the relationship between HMWS proportion changes and spectral variations[42,43], 2DCOS was employed. 2DCOS is an emerging analytical tool that is harnessed to uncover the specific spectral intensity fluctuations induced by the external perturbations (i.e. the aggregation of the mAb samples). 2DCOS analysis was performed under the environment of MATLAB 2017a (Mathworks, Inc., Natick, MA, USA). The 2DCOS analysis provides synchronous and asynchronous contour maps. The details of 2DCOS interpretation are noted in the Appendix.

2.4 Results and Discussion

Rapid and accurate determination of protein aggregation remains an unmet analytical need, addressing which is crucial to the development of more efficient and inexpensive manufacturing processes of protein-based biopharmaceuticals. While previous Raman spectroscopy studies have focused on single variants of well-characterized proteins[28-30] or on understanding aggregation mechanisms of specific mAb molecules[31], this technique is yet to be applied for quantifying aggregation levels, particularly in order to assess its feasibility for detecting low levels of aggregation in ADCs. The latter forms a particularly important, yet understudied, sample cohort. Additionally, in the second part of our investigation (study (B)), we seek to evaluate the robustness of Raman spectroscopy-based multivariate algorithms in determining changes induced by thermal incubation that mimic long-term storage conditions in real life.

Study A: Quantitative determination of aggregation levels in ADC mixture samples. Based on the HP-SEC profile, peak fractionation was applied to separate major ADC product (monomer) and HMWS, and subsequently a series of mixture samples consisting of increasing amounts of HMWS was established. **Figure 2.1A** shows representative label-free Raman spectrum recorded from a clean ADC sample (i.e. without the presence of the HP-SEC mobile phase) after background fluorescence subtraction. The vibrational peaks in the range of 600-1800 cm^{-1} are identified in the spectrum corresponding to various amino acids and characteristic amide modes, such as those at 764 cm^{-1} (tryptophan), 829 cm^{-1} (tyrosine), 931 and 1069 cm^{-1} (proline), 1007 and 1031 cm^{-1} (phenylalanine), 1243 cm^{-1} (amide-III), 1342 cm^{-1} (CH deformation), 1381 cm^{-1} (CH$_3$ band), 1418 cm^{-1} (CH$_2$ bending), 1471 cm^{-1} (C=N stretching), 1566 cm^{-1} (amide-II), and

1643 cm^{-1} (amide-I)[44-46]. In **Figure 2.1B**, representative Raman spectra acquired from 0% HMWS and 50% HMWS standard samples that were obtained following peak fractionation are shown. Due to collection of the fractionated species in the HP-SEC mobile phase solution, the aggregation standard spectra were normalized to the intensity of the 983 cm^{-1} peak that is characteristic of the mobile phase background. Since the mobile phase does not exhibit any interference in the 1160-1800 cm^{-1} range (marked by the black dashed box in **Figure 2.1B**), we used this spectral subset for development of the multivariate regression models and the ensuing analysis. Several spectral differences have been found between the 0% and 50% HMWS standard samples, notably in the amide-II, CH$_2$ bending and CH deformation regions. The amide-II bond (purple dashed line), which is attributed to the secondary structure of the protein, represents C-N stretching vibrations in combination with N-H bending[46,47]. The CH$_2$ bending and CH deformation features (yellow dashed lines) are attribute to the primary structure of proteins[46]. Crucially, other subtle, but reproducible, changes correlating to different protein aggregation levels may be revealed through the application of chemometric algorithms[32]. To this end, we developed a SVM-based regression model with the goal of quantifying protein aggregation level and subsequently employed PCA to tease out aggregation-specific differences in the spectral profiles.
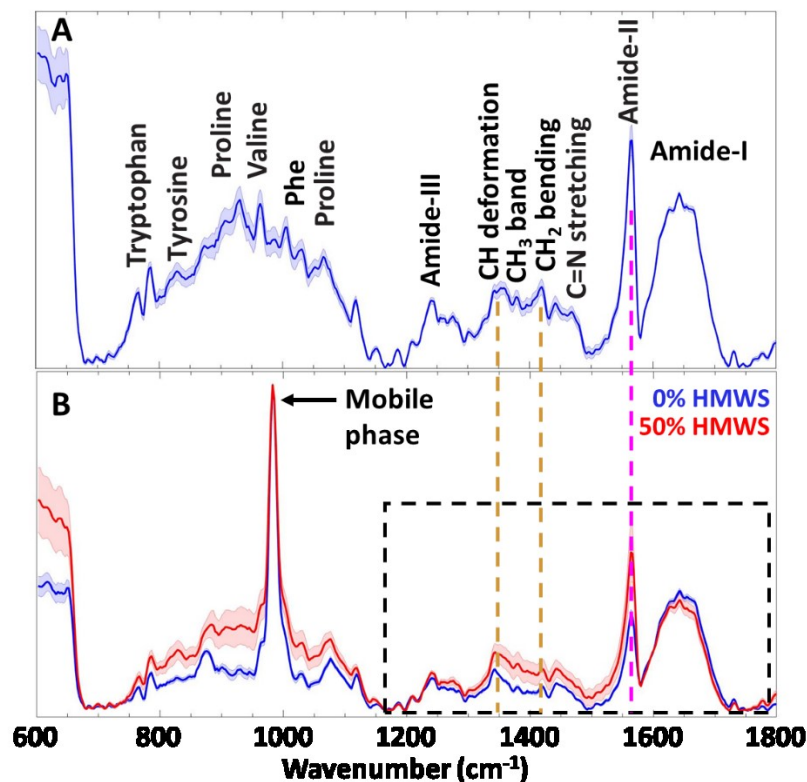
Figure 2.1 **Vibrational Raman spectra of the biologics.** (A) Label-free Raman spectra recorded from a representative mAb sample. Prominent Raman peaks are indicated in the biologic spectrum. (B) Mean spectra obtained from standard mAb samples with 0% (blue curve) and 50% (red curve) HMWS proportion after peak fractionation. The solid lines depict the mean spectrum with associated shadings representing the ±1 standard deviations (SD). The spectra were normalized to the intensity of the 983 cm^{-1} peak – characteristic of the mobile phase. The 1160-1800 cm^{-1} (black dashed box) fingerprint region was used for further multivariate model development. Spectral features of CH deformation, CH$_2$ bending and amide-II modes (from left to right) are marked by yellow and purple dashed lines.

SVM was chosen as the supervised regression technique as it is particularly well-suited to deal with data sets where the number of variables is large with undetermined linearity, two key characteristics of the recorded vibrational spectra[48,49]. **Figure 2.2A** shows the boxplot of the leave-many-out cross-validation results when the SVM-derived model is applied to the acquired spectral dataset. The root-mean-square error (RMSE) was computed to be 1.8% with 100 iterations of equal test and training spectral dataset division.

Figure 2.2B presents the boxplot of the blinded prediction results where separate samples with aggregation levels unknown to the developed regression model were used. The developed model maintains its accuracy for estimating HMWS content in the range of 0-30% with RMSE of *ca.* 4.5%. What is most impressive is that RS can make accurate and precise predictions in near real-time, with minimal processing at low aggregation levels of the mAbs, which present singular quantification challenges for DLS.

However, the prediction error in the blinded tests increased when dealing with samples containing higher relative amounts of HMWS (40% or 50%, **Figure 2.2B** red dashed boxes). In particular, test samples with 40% and 50% HMWS content were spectroscopically estimated to have, on average, only ca. 28% and 38% HMWS proportions, respectively. (In contrast, the cross-validation results do not show significant deviation at these high aggregation levels.) A possible reason for the deviation may be an outcome of the larger number of samples with low aggregation levels in the training data that may artificially skew the regression weights in favor of the latter. It is notable that the underestimation at these levels is likely not significant in the context of Raman-based quality control of the mAb products, as the overall assessment that these samples harbor large amounts of HMWS is correct and would offer the desired input to the analytical process. Nevertheless, to reveal the potential basis for these deviations at high levels of aggregation, we resorted to principal component analysis.
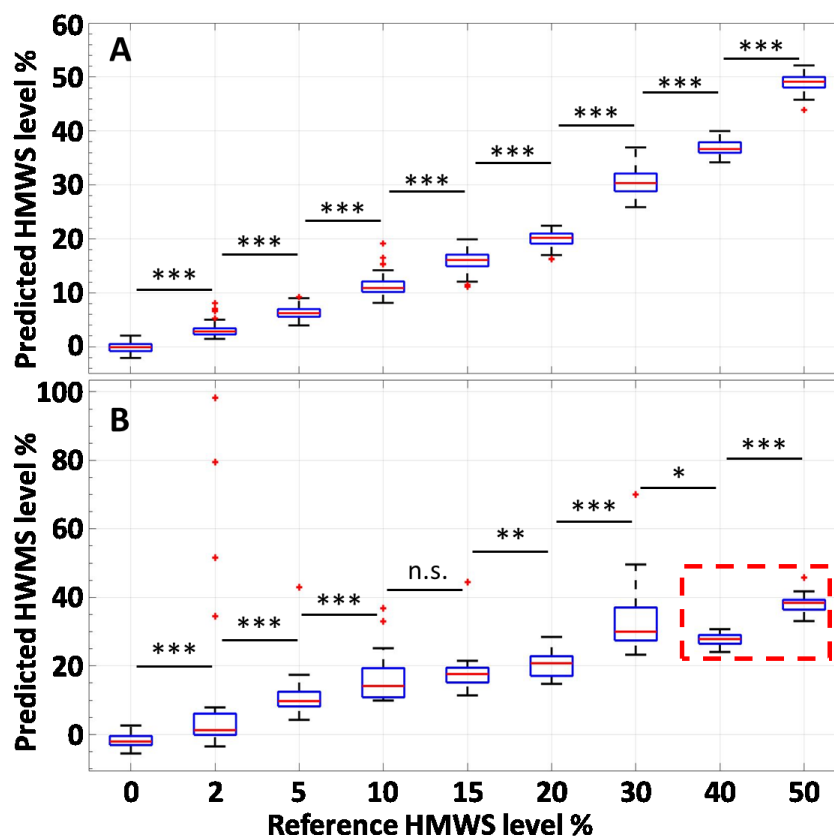
Figure 2.2 **Aggregation level predictions with label-free Raman spectroscopy.** The median values indicate the prediction results. (A) Graphical representation of prediction results obtained by application of the SVM-derived regression algorithm in leave-many-out cross-validation routine. (B) The SVM-derived regression results of protein aggregation in blinded experiments. The red dashed box highlights the predictions for 40% and 50% aggregations. The root-mean-square error (RMSE) for cross-validation is computed to be 1.8%. The RMSE for the blinded tests (for 0-30% aggregation levels) is 4.5%. (n.s.: not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Figure 2.3A shows PC-scores based radial visualization plots for various aggregation standards. Points on the radial map represent sample spectral measurements from several different experimental batches, and the scores corresponding to the chosen loadings influence their position[41]. In comparison to the regular 2D/3D PC score plots, these radial visualization plots offer substantial advantages, notably their ability to utilize more than three PCs and to depict all of the clusters in a single plane. Evidently, distinct clustering patterns for the data points possessing similar content of HMWS species is observed.

Further inspection reveals a trending of the PC scores associated with increasing HMWS content from the left to the right of the unit circle. The only major intra-class variability is seen for the samples with sub-10% (but not 0%) HMWS content. We attribute this spread to the difficulty in monitoring low concentration of sub-visible particles that freely diffuse into/out of the laser's focal volume.

Figure 2.3B shows the corresponding loadings of PC 1 to PC 4. The major peak in these loadings is located at 1566 cm^{-1}, which is the amide-II bond representing C-N stretching vibrations in combination with N-H bending[46,47]. Our observations are consistent with findings from previous studies where the vibrational information of the amide-II peak has been used to measure protein clustering[50-53]. Additionally, in the context of separating between the samples with 10-20% and 30% HMWS content, the CH deformation and CH$_2$ bending, as seen in the PC1 loading, may play an essential role.
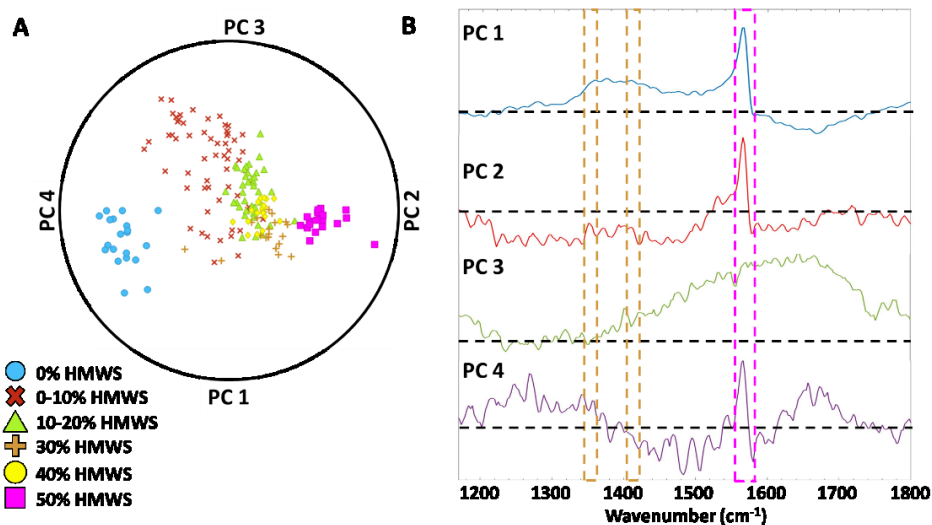
Figure 2.3 **Radial visualization plot highlights changes correlating with increasing HMWS proportion.** (A) Multidimensional radial visualization plot based on the PC scores shows the clustering behavior of six representative aggregation levels. (B) Corresponding PC 1 and 2 loadings. Black dashed lines indicate zero loading positions. Yellow and purple dashed boxes indicate (from left to right) the CH deformation, CH$_2$ bending and amide-II modes.

In order to gain a better understanding of spectral intensity fluctuation induced by changing aggregation levels, synchronous (**Figure 2.4A**) and asynchronous (**Figure 2.4B**) 2DCOS plots are presented here. In the synchronous plot shown in **Figure 2.4A**, the bands with the most obvious dynamic spectral variations reflecting aggregation changes are observed at 1342, 1381, 1418 and 1566 cm^{-1} (CH deformation, CH$_3$ band, CH$_2$ bending and amide-II, respectively). The positive cross peaks occurring between (1566, 1342) and (1566, 1418) accurately present that both amide-II and CH-related bands are changing intensity in the same direction as aggregation proportion increase. Besides positive cross peaks, a weak negative cross peak is noted around (1566, 1643) cm^{-1}, which could be assigned to amide-II / amide-I cross peak, indicating their changes are in opposite directions. In the asynchronous plot (**Figure 2.4B**), there is a major positive cross peak located between the coordinates (1566, 1243) and (1566, 1500) cm^{-1} that correspond to amide-II, amide-III, C=N stretching and CH-related bands. However, due to the absence of obvious corresponding cross peaks around (1566, 1243) and (1566, 1471) cm^{-1} in the synchronous plot, the predominant order of amide-II and amide-III or C=N stretching bands intensity variance cannot be determined at this stage. Since the corresponding amide-II and CH-related cross peaks share the same sign (positive) in both the synchronous and asynchronous plots, it is indicative of the aggregation-induced changes in amide-II intensity occurring predominantly ahead of such changes in the CH-related spectral features.
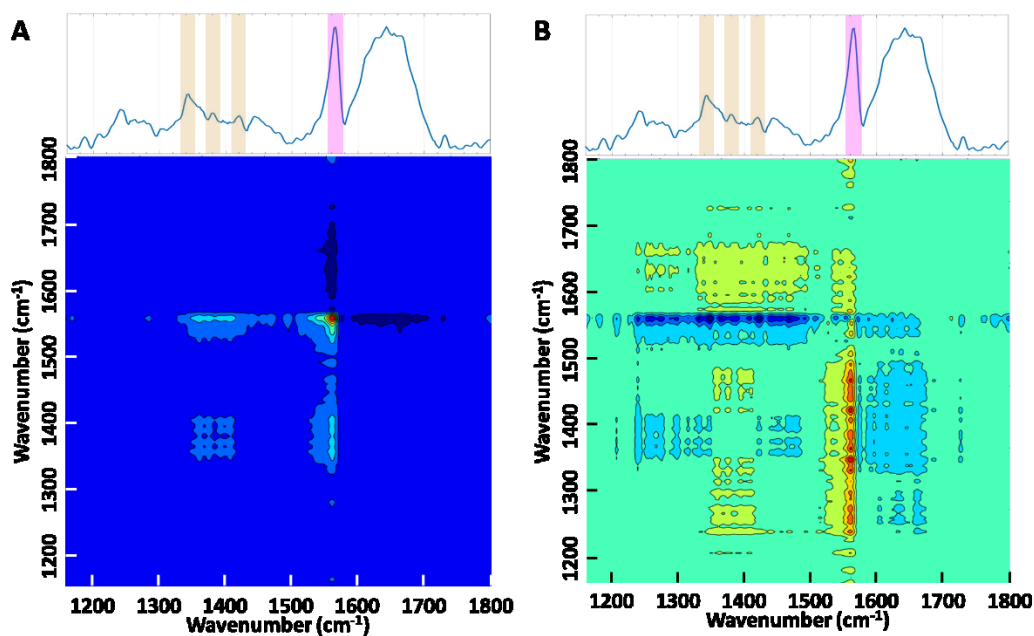
Figure 2.4 (A) Synchronous and (B) asynchronous 2DCOS plots generated from mAb aggregation-induced Raman spectral intensity variations. The shaded regions above the 2DCOS plots highlight (from left to right) the CH deformation, CH_3 band, CH_2 bending and amide-II modes in the Raman spectrum.

The inclusion of the CH and CH_2 peaks in accounting for this differential separation indicates that the mechanism of aggregation is driven, in part, by the attraction between hydrophobic patches of the protein molecules, involving noncovalent interactions[31]. As mentioned previously, we have not denatured the protein under investigation in this study and, as such, the protein is predominantly in its native state, resulting in the presence of surface exposed hydrophobic amino acids, which likely act as nucleation sites for aggregation to occur via hydrophobic attraction between non-polar side chains[54]. Similar conclusion can also be drawn from **Figure 2.1B**, where the largest variance in our data is observed in the amide-II region with secondary variances distributed around CH deformation and CH_2 bending modes. The amide-I region is widely studied to assess protein secondary structure[55,56]. However, due to the broad band of H-O-H bending

vibrations from water (as also reported by previous studies[57]), amide-I does not appear to be a major driver of our regression model for quantifying aggregation levels.

Study B: Quantification of aggregation levels of mAbs subjected to thermal stress. Here, we sought to investigate the feasibility of the previously developed regression model in quantifying differential aggregation levels of the ADC sample, pre- and post-thermal treatment. This experiment is of significance in evaluating the suitability of RS in monitoring protein degradation over the long-term and, specifically, in understanding the robustness of our algorithm to non-analyte specific buffer variations. The SVM model, developed in the aforementioned cross-validation study, was used without alteration to predict aggregate content in the pre- and post-stress ADC samples, which (unlike the samples in study (A)) do not consist of the HP-SEC mobile phase.
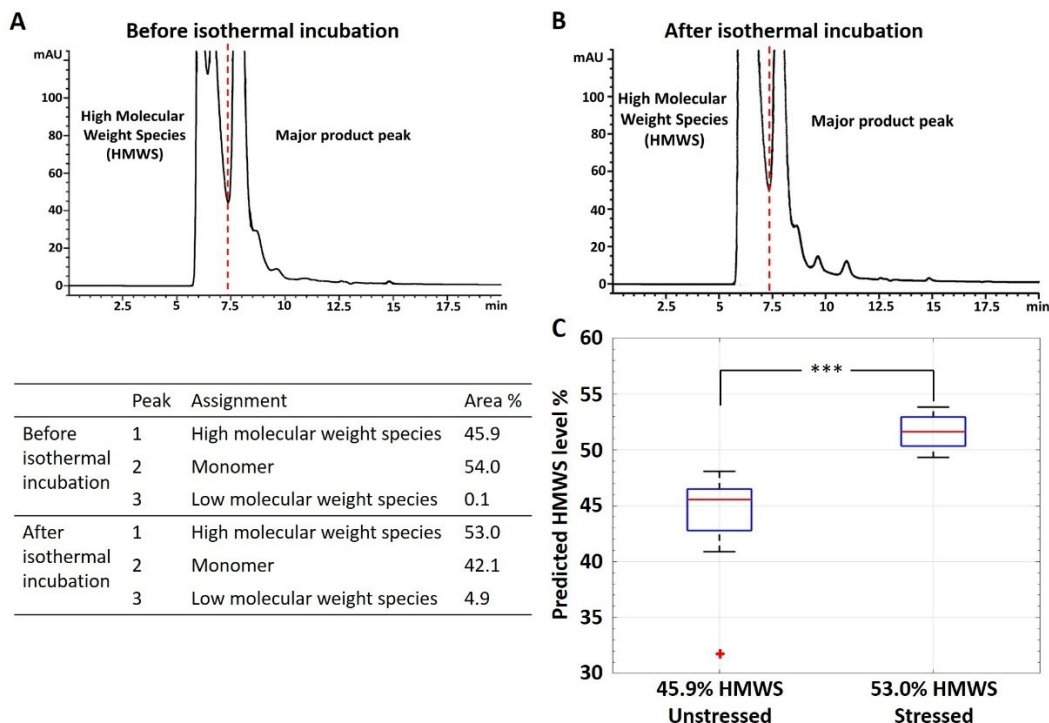
Figure 2.5 Size exclusion chromatography characterization of mAb specimen: (A) before isothermal incubation; and (B) after one-month isothermal incubation. The monomer/HMWS separation points are indicated by the red dashed lines. The included

34

tables present the HP-SEC peak assignments and area coverage of each component. (C) Aggregation level predictions on pre-/post- isothermal incubation samples with label-free Raman spectroscopy. Independent prediction results were obtained by applying the SVM-based regression model from study A. The SVM-derived model estimated the HMWS component in the pre- and post- isothermal incubation samples to be 45.7% and 51.8%, respectively. The RMSE of prediction on unstressed/stressed protein aggregation is 1.4%. (n.s.: not significant, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

Our unstressed ADC sample was an in-process intermediate from the purification process and was determined by HP-SEC to consist of 54% monomer and 45.9% HMWS, respectively (**Figure 2.5A**). After the one-month isothermal incubation period, the stressed sample displayed an increase in protein aggregation with 53% components now recognized as HMWS (**Figure 2.5B**). The reduction of the peak of the major product and the corresponding gain in the HMWS peak underscores the expected conversion of monomer species into larger than monomer HMWS. **Figure 2.5C** presents the boxplot of the RS-based predictions of the HMWS content in the unstressed and stressed samples. The SVM-based regression model estimates the HMWS content in the unstressed and stressed samples to be 45.7% and 51.8%, respectively. The computed RMSE is *ca.* 1.4%, which is comparable to the prediction errors observed in study (A). This verifies that the mobile phase background signal does not affect the regression model performance, and that our SVM derived regression algorithms are capable of accurately predicting mAb aggregation levels, despite the potential sample-to-sample variations. Our ability to distinguish between samples with relatively similar HMWS content paves the way for further quantitative studies to interrogate varying aggregation behavior of specific mAbs as well as the differential impact of other crucial physiochemical conditions on generation of protein particles.

2.5 Conclusion

Limitations in detecting and quantifying protein particles complicates manufacturing and quality control for monoclonal antibody-based therapeutics. Here, we have shown that Raman spectroscopy in conjunction with support vector machine regression enables quantitative prediction of protein clustering in mAbs with high degree of accuracy and robustness. Notably, each measurement took less than three minutes including spectra acquisition and data analysis, which is much faster than HP-SEC measurements (~20 min per samples) that is extensively used for monitoring protein aggregation in the biopharmaceutical industry. The results in this proof-of-concept study should not be considered as indicative of the best classification performance that is likely to be obtainable after further optimization of the spectroscopic hardware and regression algorithm. In addition, we have uncovered the regions of spectral variances induced by protein particle formation and correlated these with the aggregation mechanisms through PCA and 2DCOS. Our present findings encourage further development of this promising technique with the goal of eventual application as a rapid, in-line aggregation monitoring tool. This has extensive implications in the rapidly emerging biopharmaceutical space especially in facilities handling production and processing of biologics, and may also benefit regulatory authorities by helping develop improved guidance parameters for manufacturing of safe and effective protein therapeutics.

2.6 Appendix

2D correlation analysis (2DCOS) interpretation. The synchronous contour map is symmetric and consists of two types of contours – autopeaks and cross peaks, which are

positioned along the diagonal and off the diagonal, respectively. The autopeaks, which are always positive by definition[42,43]43, identify the Raman bands with dynamic intensity variations. On the other hand, the cross peaks, which could be positive or negative, represent the similarities between the variations of the independent Raman bands. Positive cross peak indicates the two independent bands' intensities changing in the same direction. The negative cross peak depicts the two bands varying in opposite directions.

The asynchronous plot interpretation is based on the signs of the corresponding synchronous and asynchronous cross peaks. The rules for discerning the asynchronous cross peaks for Raman spectral analysis state that if the asynchronous cross peak located at Raman shifts (v_1, v_2) with $v_1 \neq v_2$ is of the same sign as the corresponding synchronous cross peak, spectral intensity variations at wavenumber v_1 are considered to be occurring predominantly ahead of such changes at wavenumber v_2 [42,43]43. In contrast, if the asynchronous cross peak does not have the same sign as its corresponding synchronous cross peak, the reversed sequence of spectral intensity changes is occurring with intensity changes measured at v_2 happening ahead of intensity changes observed at v_1. If a cross peak is identified on the synchronous plot without the corresponding peak on the asynchronous plot, the intensity changes of both the Raman bands occur simultaneously. Finally, if an asynchronous cross peak is observed without the corresponding presence of a related synchronous cross peak, the order of band intensity variance cannot be determined.

Support vector machines (SVM). SVM is a supervised learning model that has been widely applied in the pattern recognition and classification of the spectroscopic field[58-60]. The basic concept of SVM is risk minimization, which could be explained as following: looking for an optimal hyperplane classifier to separate different classes while the distance

(margin) between classes' boundaries is maximized. The progress of margin maximization simultaneous minimizes the empirical classification error. When the data set cannot be separated by a linear classifier, the data point is projected into higher dimensional space that is decided by the selected kernel function. Then, a linear classifier will be searched to achieve the maximum margin between classes under this higher dimensional space. In order to maintain accuracy with the existence of inseparable points (outliers) even in a higher dimensional space, the insensitive zone is employed to reduce the SVM sensitivity on these points[61,62].

If the data set is linearly separable, the equation of a linear SVM can be expressed as[63]:

$$f(x) = \sum_{i=0}^{n_{sv}} \alpha_i y_i < x_i^{sv} \cdot x > + b$$

where n_{sv} represents the number of support vectors, α_i (Lagrange multipliers) is a non-negative parameter from the data set, each x_i is described by a Lagrange multiplier and b is bias.

For the situation of non-linearly separable data, the equation could be modified as:

$$f(x) = \sum_{i=0}^{n_{sv}} \alpha_i y_i k(x_i, x) + b$$

$k(x_i, x)$ is the kernel function.
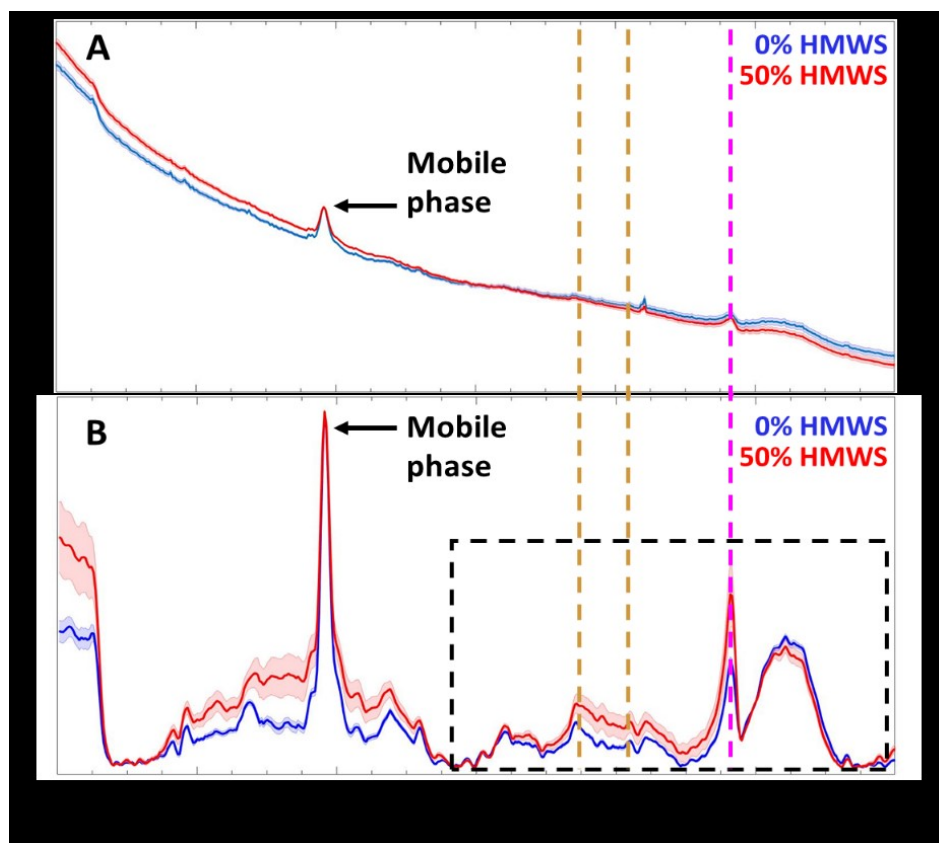
Figure S2.1 **Vibrational Raman spectra before and after background removal.** (A) Label-free Raman spectra from standard mAb samples with 0% (blue curve) and 50% (red curve) HMWS proportion before fifth order polynomial background removal. (B) Mean spectra obtained from standard mAb samples with 0% (blue curve) and 50% (red curve) HMWS proportion after polynomial background removal. The solid lines depict the mean spectrum with associated shadings representing the ±1 standard deviations (SD). Prominent Raman peaks are indicated in the biologic spectrum. The spectra were normalized to the intensity of the 983 cm^{-1} peak – characteristic of the mobile phase background. The 1160-1800 cm^{-1} (black dashed box) fingerprint region was used for further multivariate model development. Spectral features of CH deformation, CH_2 bending and amide-II modes (from left to right) are marked by yellow and purple dashed lines.

Acknowledgements

Bibliography

1. Nelson, A.L., Dhimolea, E. and Reichert, J.M., Development trends for human monoclonal antibody therapeutics, *Nat. Rev. Drug Discov.* **2010**, 9, pp. 767-774.

2. Reichert, J.M. Antibodies to watch in 2017, *MAbs* **2017**, 9, pp. 167-181.

3. Grilo, A.L. and Mantalaris, A. The Increasingly Human and Profitable Monoclonal Antibody Market, *Trends Biotechnol.* **2018**, 37, pp. 9-16.

4. Maggon, K. Monoclonal antibody "gold rush", *Curr. Med. Chem.* **2007**, 14, pp. 1978-1987.

5. De Baets, G., Schymkowitz, J. and Rousseau, F. Predicting aggregation-prone sequences in proteins, *Essays Biochem.* **2014**, 56, pp. 41-52.

6. Singla, A., Bansal, R., Joshi, V. and Rathore, A.S. Aggregation kinetics for IgG1-based monoclonal antibody therapeutics, *AAPS J.* **2016**, 18, pp. 689-702.

7. Rosenberg, A.S. Effects of protein aggregates: an immunologic perspective, *AAPS J.* **2006**, 8, pp. E501-E507.

8. Ratanji, K.D., Dearman, R.J., Kimber, I., Thorpe, R., Wadhwa, M. and Derrick, J.P. Editor's Highlight: Subvisible Aggregates of Immunogenic Proteins Promote a Th1-Type Response, *Toxicol. Sci.* **2016**, 153, pp. 258-270.

9. Sluzky, V., Shahrokh, Z., Stratton, P.A., Eberlein G.A. and Wang, Y.J. Chromatographic methods for quantitation of native, denatured and aggregated bFGF in solution formulations, *Pharm. Res.* **1994**, 11, pp. 485-490.

10. Uversky, V.N. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule, *Biochemistry* **1993**, 32, pp. 13288-13298.

11. Hartmann, W.K., Saptharishi, N., Yang, X.Y., Mitra, G. and Soman, G. Characterization and analysis of thermal denaturation of antibodies by size exclusion high-performance liquid chromatography with quadruple detection, *Anal. Biochem.* **2004**, 325, pp. 227-239.

12. Wang, W. Instability, stabilization, and formulation of liquid protein pharmaceuticals, *Int. J. Pharm.* **1999**, 185, pp. 129-188.

13. Fraunhofer, W. and Winter, G. The use of asymmetrical flow field-flow fractionation in pharmaceutics and biopharmaceutics, *Eur. J. Pharm. Biopharm.* **2004**, 58, pp. 369-383.

14. Ladokhin, A.S. Fluorescence spectroscopy in peptide and protein analysis, *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*. **2006**. https://doi.org/10.1002/9780470027318.a1611

15. Kelly, S.M., Jess, T.J. and Price, N.C. How to study proteins by circular dichroism, *Biochim. Biophys. Acta.* **2005**, 1751, pp. 119-139.

16. Joshi, V., Shivach, T., Yadav, N. and Rathore, A.S. Circular dichroism spectroscopy as a tool for monitoring aggregation in monoclonal antibody therapeutics, *Anal. Chem.* **2014**, 86, pp. 11606-11613.

17. Yu, Z., Reid, J.C. and Yang, Y.P. Utilizing dynamic light scattering as a process analytical technology for protein folding and aggregation monitoring in vaccine manufacturing, *J. Pharm. Sci.* **2013**, 102, pp. 4284-4290.

18. Kükrer, B., Filipe, V., van Duijn, E., Kasper, P.T., Vreeken, R.J., Heck, A.J. and Jiskoot, W. Mass spectrometric analysis of intact human monoclonal antibody

aggregates fractionated by size-exclusion chromatography, *Pharm. Res.* **2010**, 27, pp. 2197-2204.

19. Beck, A., Sanglier-Cianférani, S. and Van Dorsselaer, A. Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry, *Anal. Chem.* **2012**, 84, pp. 4637-4646.

20. Amin, S., Barnett, G.V., Pathak, J.A., Roberts, C.J. and Sarangapani, P.S. Protein aggregation, particle formation, characterization & rheology, *Curr. Opin. Colloid In.* **2014**, 19, pp. 438-449.

21. Tatford, O.C., Gomme, P.T. and Bertolini, J. Analytical techniques for the evaluation of liquid protein therapeutics, *Biotechnol. Appl. Biochem.* **2004**, 40, pp. 67-81.

22. Wen, Z.Q. Raman spectroscopy of protein pharmaceuticals, *J. Pharm. Sci.* **2007**, 96, pp. 2861-2878.

23. Navarra, G., Tinti, A., Leone, M., Militello, V. and Torreggiani, A. Influence of metal ions on thermal aggregation of bovine serum albumin: aggregation kinetics and structural changes, *J. Inorg. Biochem.* **2009**, 103, pp. 1729-1738.

24. Brewster, V.L., Ashton, L. and Goodacre, R. Monitoring guanidinium-induced structural changes in ribonuclease proteins using Raman spectroscopy and 2D correlation analysis, *Anal. Chem.* **2013**, 85, pp. 3570-3575.

25. Zheng, R., Zheng, X., Dong, J. and Carey, P.R. Proteins can convert to β-sheet in single crystals, *Protein Sci.* **2004**, 13, pp. 1288-1294.

26. Sane, S.U., Wong, R. and Hsu, C.C. Raman spectroscopic characterization of drying-induced structural changes in a therapeutic antibody: correlating structural changes with long-term stability, *J. Pharm. Sci.* **2004**, 93, pp. 1005-1018.

27. Wen, Z.Q., Cao, X. and Vance, A. Conformation and side chains environments of recombinant human interleukin-1 receptor antagonist (rh-IL-1ra) probed by raman, raman optical activity, and UV-resonance Raman spectroscopy, *J. Pharm. Sci.* **2008**, 97, pp. 2228-2241.

28. Zhou, C., Qi, W., Lewis, E.N. and Carpenter, J.F. Concomitant Raman spectroscopy and dynamic light scattering for characterization of therapeutic proteins at high concentrations, *Anal. Biochem.* **2015**, 472, pp. 7-20.

29. Zhou, C., Qi, W., Lewis, E.N. and Carpenter, J.F. Characterization of sizes of aggregates of insulin analogs and the conformations of the constituent protein molecules: a concomitant dynamic light scattering and Raman spectroscopy study, *J. Pharm. Sci.* **2016**, 105, pp. 551-558.

30. Lewis, E.N., Qi, W., Kidder, L.H., Amin, S., Kenyon, S.M. and Blake, S. Combined dynamic light scattering and Raman spectroscopy approach for characterizing the aggregation of therapeutic proteins, *Molecules* **2014**, 19, pp. 20888-20905.

31. Gómez de la Cuesta, R., Goodacre, R. and Ashton, L. Monitoring antibody aggregation in early drug development using Raman spectroscopy and perturbation-correlation moving windows, *Anal. Chem.* **2014**, 86, pp. 11133-11140.

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

33. Abu-Absi, N.R., Kenty, B.M., Cuellar, M.E., Borys, M.C., Sakhamuri, S., Strachan, D.J., Hausladen, M.C. and Li, Z.J. Real time monitoring of multiple parameters in

mammalian cell culture bioreactors using an in-line Raman spectroscopy probe, *Biotechnol. Bioeng.* **2011**, 108, pp. 1215-1221.

34. De Beer, T., Burggraeve, A., Fonteyne, M., Saerens, L., Remon, J.P. and Vervaet, C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes, *Int. J. Pharm.* **2011**, 417, pp. 32-47.

35. ICH Q5C. Quality of biotechnological products: stability testing of biotechnological/biological products, **1995**.

36. McConnell, A.D., Zhang, X., Macomber, J.L., Chau, B., Sheffer, J.C., Rahmanian, S., Hare, E., Spasojevic, V., Horlick, R.A., King, D.J. and Bowers, P.M. A general approach to antibody thermostabilization, *MAbs* **2014**, 6, pp. 1274-1282.

37. Genton, M.G. Classes of kernels for machine learning: a statistics perspective, *J. Mach. Learn. Res.* **2001**, 2, pp. 299-312.

38. Suykens, J.A., Van Gestel, T., Vandewalle, J. and De Moor, B. A support vector machine formulation to PCA analysis and its kernel version, *IEEE Trans. Neural. Netw.* **2003**, 14, pp. 447-450.

39. Pelckmans, K., Suykens, J.A., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B. and Vandewalle, J. LS-SVMlab: a matlab/c toolbox for least squares support vector machines, *Tutorial. KULeuven-ESAT. Leuven, Belgium*, **2002**, 142, pp. 1-2.

40. Ringnér, M. What is principal component analysis?, *Nat. Biotechnol.* **2008**, 26, pp. 303.

41. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A. and Štajdohar, M. Orange: data mining toolbox in Python, *J. Mach. Learn. Res.* **2013**, 14, pp. 2349-2353.

42. Noda, I. Two-dimensional infrared (2D IR) spectroscopy: theory and applications, *Appl. Spectrosc.* **1990**, 44, pp. 550-561.

43. Noda, I., Roy, A., Carriere, J., Sobieski, B.J., Chase, D.B. and Rabolt, J.F. Two-Dimensional Raman Correlation Spectroscopy Study of Poly[(R)-3-hydroxybutyrate-co-(R)-3-hydroxyhexanoate] Copolymers, *Appl Spectrosc.* **2017**, 71, pp. 1427-1431.

44. De Gelder, J., De Gussem, K., Vandenabeele, P. and Moens, L. Reference database of Raman spectra of biological molecules, *J. Raman Spectrosc.* **2007**, 38, pp. 1133-1147.

45. Huang, Z., Chen, X., Chen, Y., Feng, S., Chen, R., Chen, J., Dou, M. and Zeng, H. Raman spectroscopic characterization and differentiation of seminal plasma, *J. Biomed. Opt.* **2011**, 16, pp. 110501.

46. Movasaghi, Z., Rehman, S. and Rehman, I.U. Raman spectroscopy of biological tissues, *Appl. Spectrosc. Rev.* **2007**, 42, pp. 493-541.

47. Krimm, S. and Bandekar, J. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins, *Adv. Protein Chem.* **1986**, 38, pp. 181-364.

48. Pandey, R., Zhang, C., Kang, J.W., Desai, P.M., Dasari, R.R., Barman, I. and Valdez, T.A. Differential diagnosis of otitis media with effusion using label-free Raman spectroscopy: A pilot study, *J. Biophotonics.* **2018**, 11, pp. e201700259.

49. Zhang, C., Winnard Jr, P.T., Dasari, S., Kominsky, S.L., Doucet, M., Jayaraman, S., Raman, V. and Barman, I. Label-free Raman spectroscopy provides early determination and precise localization of breast cancer-colonized bone alterations, *Chem. Sci.* **2018**, 9, pp. 743-753.

50. Boulet-Audet, M., Byrne, B. and Kazarian, S.G. High-throughput thermal stability analysis of a monoclonal antibody by attenuated total reflection FT-IR spectroscopic imaging, *Anal. Chem.* **2014**, 86, pp. 9786-9793.

51. Lin, G.L., Pathak, J.A., Kim, D.H., Carlson, M., Riguero, V., Kim, Y.J., Buff, J.S. and Fuller, G.G. Interfacial dilatational deformation accelerates particle formation in monoclonal antibody solutions, *Soft Matter* **2016**, 12, pp. 3293-3302.

52. Schüle, S. Frieß, W., Bechtold-Peters, K. and Garidel, P. Conformational analysis of protein secondary structure during spray-drying of antibody/mannitol formulations, *Eur. J. Pharm. Biopharm.* **2007**, 65, pp. 1-9.

53. Ami, D., Lavatelli, F., Rognoni, P., Palladini, G., Raimondi, S., Giorgetti, S., Monti, L., Doglia, S.M., Natalello, A. and Merlini, G. In situ characterization of protein aggregates in human tissues affected by light chain amyloidosis: a FTIR microspectroscopy study, *Sci. Rep.* **2016**, 6, pp. 29096.

54. Roberts, C.J. Therapeutic protein aggregation: mechanisms, design, and control, *Trends Biotechnol.* **2014**, 32, pp. 372-380.

55. Wen, Z.Q. Raman spectroscopy of protein pharmaceuticals, *J. Pharm. Sci.* **2007**, 96, pp. 2861-2878.

56. Ye, S., Li, H., Yang, W. and Luo, Y. Accurate determination of interfacial protein secondary structure by combining interfacial-sensitive amide I and amide III spectral signals, *J. Am. Chem. Soc.* **2014**, 136, pp. 1206-1209.

57. Ashton, L., Johannessen, C. and Goodacre, R. The importance of protonation in the investigation of protein phosphorylation using Raman spectroscopy and Raman optical activity, *Anal. Chem.* **2011**, 83, pp. 7978-7983.

58. Dingari, N.C., Barman, I., Myakalwar, A.K., Tewari, S.P. and Kumar Gundawar, M. Incorporation of support vector machines in the LIBS toolbox for sensitive and robust classification amidst unexpected sample and system variability. *Analytical chemistry*, **2012**, 84, pp. 2686-2694.

59. Yan, B., Li, B., Wen, Z., Luo, X., Xue, L. and Li, L. Label-free blood serum detection by using surface-enhanced Raman spectroscopy and support vector machine for the preoperative diagnosis of parotid gland tumors. *BMC cancer*, **2015**, 15, pp. 650.

60. Sikirzhytskaya, A., Sikirzhytski, V. and Lednev, I.K. Determining gender by Raman spectroscopy of a bloodstain. *Analytical chemistry*, **2017**, 89, pp. 1486-1492.

61. Khan, S., Ullah, R., Khan, A., Wahab, N., Bilal, M. and Ahmed, M. Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM). *Biomedical optics express*, **2016**, 7, pp. 2249-2256.

62. Dong, R., Weng, S., Yang, L. and Liu, J. Detection and direct readout of drugs in human urine using dynamic surface-enhanced Raman spectroscopy and support vector machines. *Analytical chemistry*, **2015**, 87, pp. 2937-2944.

63. Pierna, J.F., Baeten, V., Renier, A.M., Cogdill, R.P. and Dardenne, P. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics: A Journal of the Chemometrics Society*, **2004**, 18, pp. 341-349.

Chapter 3. Composite-scattering plasmonic nanoprobes for label-free, quantitative biomolecular sensing

3.1 Abstract

Biosensing based on localized surface plasmon resonance (LSPR) relies on concentrating light to a nanometeric spot and leads to a highly enhanced electromagnetic field near a metal nanostructure. Here, we present a design of plasmonic nanostructures based on rationally structured metal-dielectric combinations, which we call composite scattering probes (CSP), to generate an integrated multi-modal biosensing platform featuring LSPR and surface-enhanced Raman scattering (SERS) measurements. Specifically, we propose CSP configurations that have several prominent resonance peaks enabling higher tunability and sensitivity for self-referenced multiplexed analyte sensing. Using electron-beam evaporation and thermal de-wetting, we have fabricated large area, uniform, and tunable CSP, which are suitable for label-free LSPR and SERS measurements. The CSP prototypes were used to demonstrate refractive index sensing and molecular analysis using albumin as a model analyte. By using partial least squares on recorded absorption profiles, differentiation of subtle changes in refractive index (as low as 0.001) in the CSP milieu was demonstrated. Additionally, CSP facilitates complementary untargeted plasmon-enhanced Raman measurements from the sample's compositional contributors. With further refinement, we envision that our method may lead to a sensitive,

versatile and tunable platform for quantitative concentration determination and molecular fingerprinting, particularly where limited *a priori* information of the sample is available.

3.2 Introduction

Biosensors offering fast and accurate detection of biomarkers can serve as important tools for point-of-care medical diagnostics[1,2]. Tagging the biomarkers with fluorescent species and detecting them using fluorescence microscopy or assay-based techniques are common approaches[3-5]. However, addition of exogenous contrast agents may perturb the native environment in cellular or biomarker sensing and intrinsic contrast imaging remains a highly desired commodity. In this milieu, nanoplasmonic sensors provide a highly sensitive and label-free method for detection of target molecules[6-9]. For instance, ligand-receptor binding events are regularly monitored by detecting variations in surface plasmon resonance (SPR)[10,11], which involves the collective oscillation of conduction electrons that propagates along a metal-dielectric interface. The excitation of such waves is highly sensitive to the variation of refractive index (RI) at the metal surface due to the generation of strong evanescent waves[10]. The decay length of these evanescent waves is typically hundreds of nanometers, and biomolecular binding events in this range can be detected in real time by tracking the variation in SPR signal[12]. Sensors based on SPR are therefore suitable for tracking bulk RI change. Localized surface plasmon resonance (LSPR), on the other hand, confines the plasmon resonance within metallic nanostructures as the near field decays rapidly near the nanostructures[10]. Thus, this technique operates with low sensing volume and is sensitive to local RI change, making it suitable for integration with lab-on-a-chip devices[13-15].

There have been a range of LSPR-based RI sensors reported in the past[6]. Various plasmonic nanostructures have been fabricated by both chemical[16] and physical[16,17] methods for LSPR-based sensing. For example, a gold mushroom array was fabricated by interference lithography for detection of cytochrome *c* and alpha-fetoprotein[16-18], gold nanoholes and discs were fabricated by nanosphere lithography[16-19], and gold-coated silica nanospheres were realized by a chemical method[16-20]. There are also reports in the literature of various shapes of nanoparticles like gold[21] and silver[21-23] nanospeheres, gold stars[24,25], silver cubes[26], gold pyramids[27], gold rods[27,28] and gold nanorice[29], which have been used for LSPR sensing with varying degrees of success. Theoretical predictions of dielectric sensing based on asymmetric resonances (i.e., Fano resonance[8] and quadrupolar resonance[30]) also indicated the capability for ultrasensitive transduction of biomolecular species.

Classically, most of the reported LSPR sensors have used gold nanoparticles/nanostructures with a single resonance peak around 500 nm for tracking the change in surrounding RI. Here, we report an alternate nanostructure design for generating plasmonic hotspots by using a combination of gold, silver and a suitable dielectric for sensing the RI of bioanalytes. In particular, the thin dielectric layer is used for separating the two metallic layers, which helps to prevent any kind of charge transfer between the latter and, crucially, gives rise to two distinct resonance peaks. This allows for more robust detection of RI changes in the media by concomitant computation of the shift in the two resonance peaks that facilitates ratiometric determination. Such a sensing technique enables a self-referenced platform[31,32] for robust detection of analytes, effectively dealing with a situation for which changes in a single peak may be influenced by non-analyte-

specific variations[33]. Moreover, our design has the added advantage of not being subject to fouling, as the reference feature (stemming from the gold nanosphere) is isolated from the environmentally-sensitive feature (that arises from the silver nanoparticles) by the physical presence of the dielectric layer. Additionally, as reported earlier[34-36], the advantage of having two distinct peaks is that it increases the dynamic range for sensing, since the sensitivities to the surrounding environment demonstrated by the two peaks are distinct in the two RI ranges. Moreover, the sensitivity and the dynamic range can be tailored by changing the metal or dielectric thicknesses as well as by simply heating the substrate.

Despite its sensitivity to the analyte-induced changes in RI, label-free LSPR sensing does not provide any information on the molecular composition. Raman spectroscopy, however, provides a unique molecular fingerprint by interrogating the vibrational transitions, offering multiplexed detection of biomolecules without necessitating the addition of any contrast agents[37,38]. The presence of plasmonic nanostructures near the analyte provides an electromagnetic enhancement of the Raman signal, thus, enabling detection at low concentrations – a technique commonly known as surface-enhanced Raman spectroscopy (SERS)[38]. Indeed, plasmon-enhanced assays have, in recent years, surpassed the detection limits of conventional fluorescence-based techniques[39,40].

We reason that a combination of LSPR and SERS measurements, which has only been recently explored in the literature[41-43] for molecular diagnostics, can be readily obtained by leveraging the unique properties of our gold-silver-dielectric nanostructures. The ability to use the same platform with no modifications for acquiring complementary pieces of information is particularly attractive. SERS and LSPR measurements offer distinct insights into the biomolecular environment. SERS, which relies on inelastic scattering of light,

provides molecular fingerprinting capabilities, owing to the wealth of information encoded in the vibrational modes of the Raman spectra. On the other hand, LSPR measurements inform on the elastic scattering and absorption events that is manifested in the refractive index of the surrounding. Coupled with its label-free nature, we envision that the molecular specificity and multiplexing capability of this dual-modal sensing construct, which we term as composite scattering probes (CSP), will facilitate the screening of biomarkers in body fluids for diagnosis and monitoring of therapy response. In this report, we demonstrate both detection capabilities by using bovine serum albumin (BSA) as a model analyte for understanding the performance metrics of our nanostructured platform. BSA is a widely used analogue of human albumin, an important circulating protein in the blood of vertebrates, whose detection represents an diagnostic challenge in itself[44].

3.3 Experimental section

Absorption spectra measurement and refractive index sensing: A Lambda 950 UV/VIS/NIR spectrometer (PerkinElmer, Waltham, MA, USA) was used to measure absorption spectra from substrates. The measured wavelengths range from 320 nm to 800 nm, covering the entire visible range. The data is collected at an interval of 2 nm for the optical characterization and 1 nm for the RI sensing experiments. The absorption spectra are measured by considering a glass coverslip as the reference. For the RI sensing experiments, the fabricated substrates, BSA solution and reference coverslip were put into quartz cuvettes that were placed in the measurement and reference light paths.

SERS measurement: Raman spectra were acquired by using an XploRA PLUS Raman microscope (HORIBA Instruments Inc., Edison, NJ, USA). Sample excitation was

achieved by using a 532 nm diode laser. The laser beam was focused on the sample through a 10X objective (MPlan N, Olympus, Center Valley, PA, USA). The backscattered Raman signals were collected by the same objective and through a notch filter to remove Raleigh elastically scattered light, and the scattered spectra were collected using a thermoelectrically cooled CCD camera (1024X256-OE Syncerity, HORIBA Instruments Inc., Edison, NJ, USA). Laser intensity at the sample was kept constant at 35 mW for all the measurements. The exposure time for each measurement was 2 seconds with 5 times accumulation. The average SERS spectra, along with the standard deviations, are plotted.

Data analysis: The collected absorption and Raman spectra were imported into a MATLAB 2018a (Mathworks, Inc., Natick, MA, USA) environment for further analysis. For RI sensing, in order to illustrate the capability to provide quantitative measurements of our substrates, hierarchical cluster analysis (HCA) and partial least squares (PLS) regression were employed. HCA is an algorithm which could group similar patterns into clusters based on Euclidean distance and average linkage method[45]. Couple dendrograms were displayed to reveal the hierarchy of clustering among samples with various RI. PLS is a widely used multiple linear regression model in quantitative spectral analyses[46]. PLS was used to establish the fundamental relations between absorption spectra and surrounding RI. For Raman measurements, spectra were processed to remove interference from cosmic rays. The spectra were restricted to the 200 cm^{-1} to 1800 cm^{-1} region and subjected to a fifth order best-fit polynomial-based baseline removal. All spectra were smoothed by using the 7 orders and 15 points Savitzky-Golay function[46,47] and normalized at the silicon background peak at 520 cm^{-1}. Limit of detection (LOD) of LSPR is calculated,

based on the IUPAC definition[48] from the best fit-line obtained between predicted concentrations and reference concentrations. The corresponding equation is:

$$LOD\ (mM) = 3\frac{S_{y/x}}{slope}\ \ where\ S_{y/x} = [\frac{\Sigma(\hat{c}_i - c_i)^2}{N - 2}]^{1/2}$$

Where $S_{y/x}$ is the standard deviation of the residuals and is a measure of the average deviation of the predicted values from the regression line. The LOD of SERS sensing was defined as the detectable signals from the lowest analyte concentration with a signal-to-noise ratio greater than 3 [49].

Simulation: A commercial finite element-based tool (COMSOL Inc., Burlington, MA, USA) was used for the numerical modeling. The nanostructures were in a square lattice and periodic boundary condition was applied in the X and Y directions of the simulation model. Through an input port in the Z direction, the nanostructures were excited. Perfectly matched layer (PML) boundary condition was applied at the input and the output port to simulate an infinite space and prevent reflection at the boundary. A dynamic tetrahedral meshing of maximum element size of 10 nm for the nanostructure and $\lambda/6$ for the rest of the space were chosen with a minimum element size of 0.1 nm. The fault tolerance was varied to check the convergence of the code. Frequency-dependent gold and silver dielectric constants from Palik[50] were employed for the simulation. The relative permittivity of silica was considered to be constant at 3.9 over the entire visible wavelength.

3.4 Results and Discussion

Fabrication of composite scattering probes (CSP). To fabricate the large-area tailored nanostructures (**Figure 3.1**), we combined physical vapor deposition with thermal annealing. Specifically, we designed and fabricated different configurations of metal-

dielectric combinations to generate a sensitive and robust composite scattering probe; two such designs are shown and analyzed here to assess their relative merits for sensing. The details of the fabrication steps and their corresponding SEM images are shown in **Figure 3.1**.

As a first step, a thin film of gold (5.5 nm) was deposited on a clean glass coverslip and a silicon wafer by electron-beam evaporation. The thin film was annealed in a muffle furnace for 6 hours at a temperature of 550^0C. Thermal de-wetting of the gold film led to the formation of near-spherical nanoparticles of size 40 nm ($\pm$10 nm). This provided a quick and simple method of achieving near-spherical nanoparticles over a large area. **Figure 3.1B** shows the scanning electron micrograph (SEM) of gold nanoparticles formed on silicon over a large area using the thermal de-wetting process. By changing the thickness of the deposited film as well as varying the annealing temperature and time, the particle size can be altered allowing the plasmon resonance of the gold spheres to be tuned in the visible range[51].

In order to decorate the gold nanoparticles with silver nanoparticles, we adopted two different techniques. In the first case, we deposited a silica layer of 10 nm on the gold nanoparticle-followed by a thin film of silver (4 nm) in a vacuum chamber with an electron-beam evaporator. The substrates were, then, annealed on a hot plate at 250^0C for 20 minutes to modify the silver islands into near-spherical shape. **Figure 3.1C** and **D** show the SEM of silver nanoparticles decorated gold nanoparticles (SNPG) before and after annealing on the hotplate, respectively. Due to low thickness, the silver film was discontinuous before annealing as seen in the SEM image in **Figure 3.1C**. Annealing allows these irregular silver islands to transform into more regular near-spherical shapes.

In the second case, we mounted the substrate with the gold nanoparticles at an oblique angle with respect to the evaporant flux (87^0 to the horizontal) in an electron-beam evaporator for subsequent deposition of 10 nm of silica and 10 nm of silver. Deposition at an extreme angle, known as oblique angle deposition (OAD)[52], ensured that the evaporant was deposited preferentially on the gold nanoparticles, and no deposition took place in the geometric shadow region of the nanoparticles. While rarely used in this context of generating biosensors[53], this technique has been shown to be suitable for fabricating various 3D porous hybrid nanostructures with options for easy tunability of the plasmon resonance by variation of shape, size, material etc. of the nanostructures[54-56]. The SEM image of silica and silver thin film-decorated gold nanoparticles (STG) is shown in **Figure 3.1E**. The zoomed-in versions (**Figure 3.1D** and **3.1E** insets) serves to highlight the structural differences between SNPG and STG. In case of SNPG, the blue dots (false color) highlight the several silver nanoparticles on top of the gold nanoparticle (represented by the yellow false color). Some of the silver nanoparticles get deposited directly on the glass/silicon substrate as well. In contrast, the STG has a blob of silver (and silica - not visible here) on the gold nanoparticle and almost no depositions on the glass/silicon substrate, indicating a perfect shadowing during deposition. The scattering-type scanning near-field optical microscopy (s-SNOM) image of one of the nanostructures on silicon wafer (SNPG) with high spatial resolution is shown in **Figure 3.1F**. A 5 μm laser wavelength, which is far from the plasmon resonance, was used for the scan to record the relative reflected image of the nanostructures.
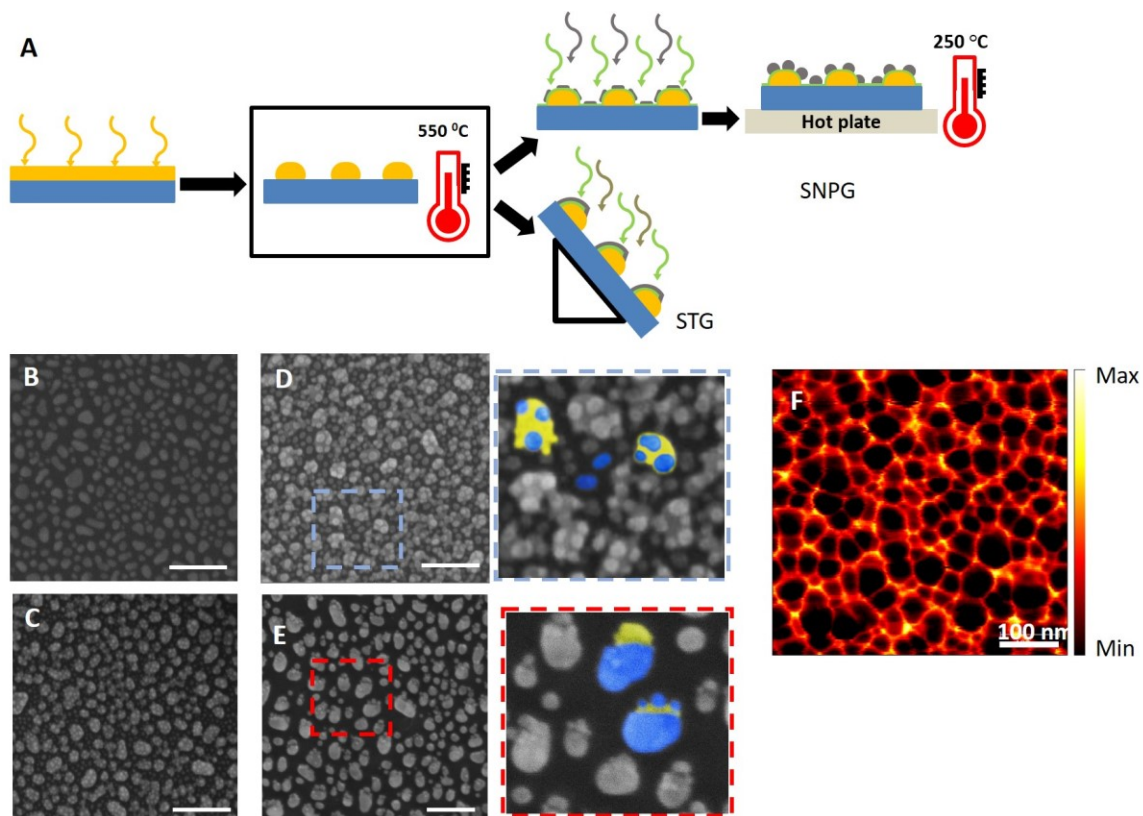
Figure 3.1 Fabrication steps: (A) Scheme of fabrication, process flow from left to right. Deposition of a gold thin film (yellow) on a glass coverslip/Si wafer using electron-beam evaporation. Thermal annealing led to de-wetting of Au film and rendered the Au islands to a near-spherical shape. For SNPG configuration, 10 nm of silica (green) and 4 nm of Ag (gray) were subsequently deposited. This was followed by annealing on a hotplate to convert the Ag islands into Ag nanoparticles. For the STG samples, OAD of 10 nm of silica and 10 nm of Ag on the Au *nanoparticles* by placing the substrates at an oblique angle relative to the evaporant flux (87^0 to the horizontal). (B) Scanning Electron Micrograph (SEM) image of the Au spheres after deposition and thermal annealing. (C) SEM image of the SNPG sample with 10 nm silica and 4 nm Ag. Since the thickness of Ag film is less, the thin film is not continuous. (D) SEM image of the SNPG sample after 20 min of annealing on a hotplate. The inset shows the zoomed-in version marked by the blue dashed box. The gold (yellow) and the silver (blue) regions of the SNPG structure are highlighted in false color. (E) SEM image of the STG sample. The inset shows the zoomed-in version marked by the red dashed box. The gold (yellow) and the silver (blue) regions of the STG structure are indicated by the false color representation. The scale bars for all the SEM images marked in white are 200 nm. (F) s-SNOM image of the annealed SNPG sample; although the silver nanoparticles cannot be resolved, the gold *nanoparticles* are readily identified (s-SNOM image courtesy: Bruker Nano Surfaces Division).

Optical characterization of CSP. We simulated optical characteristics of the silver-

decorated gold nanoparticles using a finite element method (FEM) to predict the resonance

wavelengths for the different configurations (simulation details are provided in the Materials & Methods section). Briefly, the FEM calculation was done using commercial software (COMSOL, Inc., Burlington, MA). A periodic boundary condition was considered with infinite periodicity in the X and Y direction. The perfectly matched layer (PML) boundary condition was considered in the Z direction to simulate an infinite space. A schematic diagram of the numerical model is shown in **Figure 3.2A** with the direction of propagation and polarization marked in the figure. The nanostructures were arranged in a square lattice with an edge-to-edge spacing of 50 nm. The various structures used in the numerical model are shown in **Figure 3.2B**. A gold sphere of 40 nm diameter was considered to be coated with a 10 nm silica layer. Three different sizes of silver nanoparticles of diameter 10 nm, 20 nm and 30 nm were considered to simulate different conditions of experimentally deposited silver, as can be noted in the SEM images (**Figure 3.1C, D**). The STG was simulated by considering a silica and silver thin film covering half the gold sphere mimicking the actual structures produced.

The simulated absorption spectra for the different configurations are plotted in **Figure 3.2C**. In the case of SNPG, for the silver nanoparticle of size 10 nm, the absorption cross section peak in the lower wavelength region is barely distinguishable in the spectra. For the bigger sizes of silver nanoparticles, there are two distinct peaks visible. The silica layer aids in separating the two metals to prevent hot-electron/charge transfer[57,58] between the two metals, and two distinct resonance peaks are therefore available for sensing. Due to the presence of silica and silver, the absorption peak around 560 nm is red shifted from the resonance of the gold spheres at 510 nm. The peak due to silver nanoparticles occurs around 380 nm. In the case of the STG configuration, the overall absorption cross section increased

due to the presence of the Ag thin film - with no distinct peak appearing for the silver thin film in the visible region of the electromagnetic spectrum.

For experimental optical characterization of the fabricated nanostructures, the optical absorption spectra for the different nanostructure configurations was measured using a UV-Visible spectrometer and results are plotted in **Figure 3.2D**. For the SNPG substrate with 4 nm silver coating, two distinct resonance peaks are visible, one around 580 nm, which corresponds to the plasmon resonance of the gold nanoparticle modified with silica and silver nanoparticles/islands, and the other one around 440 nm corresponding to the resonance of silver nanoparticles. Once annealed at a temperature of 250^0C, which is below the melting temperature of gold, the silver film/islands get modified into silver nanoparticles. The modification in their shape is reflected in the optical absorption spectra in **Figure 3.2D**, where the LSPR peak corresponding to silver around 440 nm diminishes for 4 nm SNPG samples. This is because on heating, the smaller silver islands become larger particles, leading to broadening and red shift of the plasmon peak[59]. Thus, the peak merges (by red-shifting) with the resonance peak of the gold and appears as a smaller shoulder near the 580 nm feature in the optical absorption spectra. The deviation from our numerical results indicates that the thermal annealing of silver in our structures is not dominated by change in size alone but can be attributed to a combination of other processes including but not limited to film oxidation and increased interparticle distance. The broadening of this silver peak due to such factors has been demonstrated previously[60], though the precise mechanism of such deviation in our case needs further investigation. Moreover, due to heating, the peak around 580 nm for the 4 nm-SNPG sample gets red shifted by about 8 nm, which is related to the further change in the dielectric environment

of the silica-coated gold spheres due to modification of the silver nanoparticles. Furthermore, the SNPG substrate shows excellent tunability. Both the resonance peaks of the SNPG sample can be easily tuned (to match the excitation for maximum enhancement) by changing the thickness of the silica layer/silver thin film or by varying the annealing time. The shift in resonance peak due to change in thickness of silver layer and annealing time is available in the Supplementary Information (**Figure S3.1** & **S3.2**).
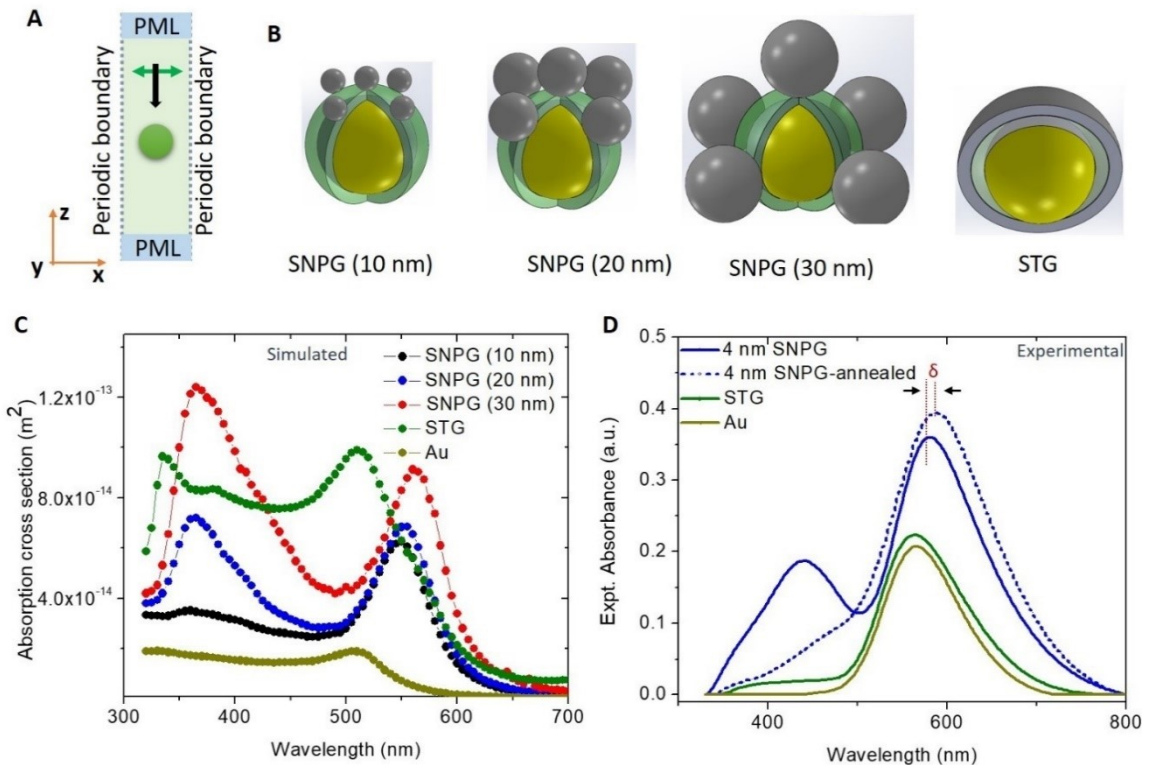
Figure 3.2 **Optical characterization of the fabricated structures, numerical and experimental**. (A) FEM 3D model with periodic boundary in the X and Y direction implying the nanostructure (represented by the orange sphere) repeats periodically in the X and Y direction infinitely. The electromagnetic field propagates along the Z direction and is linearly polarized along the X direction represented by the black and the green arrow, respectively. (B) Schematic showing various nanostructure configurations along with their respective labels (which have been used throughout this report). The yellow sphere represents the gold sphere, which is coated with silica (represented by semi-transparent green). The silver nanostructures are in gray color. (C) Simulated absorption spectra for different numerical configurations are plotted along with absorption spectra of a gold sphere. (D) Experimental absorption spectra of various fabricated samples – 4nm SNPG, 4 nm SNPG after annealing, STG and gold islands. δ signifies the shift in peak position of the SNPG sample due to annealing.

The STG sample, however, has a single resonance peak in its absorption spectra at 564 nm. This may be attributed to the dimension of the silver film being too thin to sustain a localized plasmon resonance. Moreover, in the case of the STG sample, the silver does not coat the gold nanoparticle entirely and forms an island on the gold nanoparticle. As expected, the simulated SNPG models are consistent with the pre-annealed 4 nm-SNPG samples, in part, because the size ranges considered in the simulation are similar to the pre-annealed silver islands on the gold nanostructures. The slight discrepancy in the peak position may result from the variability of particle size because the fabrication process involves physical vapor deposition.

CSP-based refractive index sensing. As is well-known, LSPR peaks are sensitive to the dielectric constant of the surrounding medium[61]. For checking the sensitivity of our substrates to the local environment, we measured the LSPR peaks of the substrates in the presence of 0.1 mmol/L, 1 mmol/L and 2 mmol/L of BSA in aqueous media. The RI of 0.1 mmol/L BSA was measured to be 1.336 as compared to 1.335 for water (measured at a temperature of 4^0C; measurement details are discussed in the Supplementary Information and plot of BSA concentration vs. RI is provided in **Figure S3.3A**). Using a UV-Visible spectrometer, we first measured the absorption spectra in air by placing the glass substrates in quartz cuvettes. Next, absorption spectra were measured by filling the cuvette serially with, water, 0.1 mmol/L, 1 mmol/L and 2 mmol/L BSA. **Figure 3.3A** plots the absorption spectra of the annealed SNPG sample and the STG sample for air and 0.1 mmol/L BSA. In the case of the SNPG sample, it is notable that, with 0.1 mmol/L BSA, a distinct peak appears near 450 nm. The latter is the approximate position of the pre-annealed feature in ambient media, which had diminished when the substrate was annealed. The peak around

580 nm shifts by 10 nm for the SNPG sample with higher RI. For the STG sample, the shift in the resonance peak due to the higher RI is 23 nm. Thus, the dominant LSPR peak is more sensitive to the surrounding media in the case of the STG sample. Nevertheless, the SNPG sample gives rise to a new spectral feature, which may be more effective in monitoring the change in RI.

To gain a better understanding of the change in absorption spectra with RI, we employed a FEM similar to the one used for the SNPG sample. A 40 nm gold sphere coated with 10 nm silica is decorated with silver nanoparticles of size 10 nm. The surrounding RI is changed from 1 to 1.3 and 1.5. **Figure 3.3B** plots the absorption spectra for these different cases. Similar to our experimental observation, at higher RI a peak appears at *ca.* 450 nm and becomes more distinct with further increase in RI. Apart from the redshift of the resonance peak, the larger size of the nanoparticle relative to the effective wavelength in presence of a higher RI media results in an increase in absorption cross-section of the silver nanoparticles.

It is worth noting that our observations for the peak corresponding to the silver nanoparticle deviate from the simulation results. While our simulations predict a red-shift in this feature with increasing bulk RI, repeated measurements establish a contrasting behavior. While it is challenging to home in on a specific mechanism for this discrepancy between experimental and numerical findings, literature reports hint at two possible explanations that focus on the effect of molecular adsorption on the nanoparticle surfaces, and alteration of the surface electronic structure of the plasmonic nanoprobes induced by the media, respectively. van Duyne and co-workers' studies detail differential blue-/red-shifting of LSPR peaks based on how analytes adsorb on nanoparticle surfaces[62], while

Muri *et al.* have postulated that the latter could explain the non-linear blue-shifting with increasing bulk RI, as they observed for gold nanorods in glycerol and sucrose solutions[63]. As these mechanisms have not been accounted for in our simulation model, uncovering the basis of this unexpected observation will be a major focus of our ongoing investigations
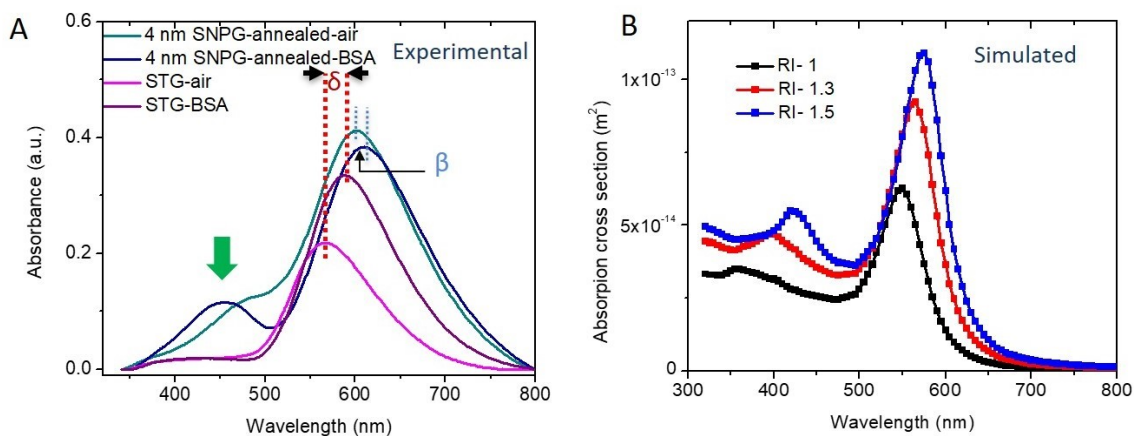
Figure 3.3 **Effect of dielectric medium on the resonance position (experiment and simulation)**. (A) Experimental absorption spectra of the annealed 4 nm SNPG and the STG samples in the presence of air and 0.1 mmol/L BSA. Green arrow indicates the new peak that appears when the SNPG sample was placed in a higher RI. δ and β indicate the shift in dominant peak positions for the STG and SNPG samples, respectively, when placed in a higher RI media. (B) (Simulated) Effect of change in dielectric medium on the absorption spectra for the model SNPG (10 nm) showing the appearance of a second peak with higher RI.

In order to establish the sensitivity of capturing small variations in RI with our CSP sensor, we leveraged chemometric analysis of the recorded spectral profiles. Elucidating subtle RI-induced changes in these profiles (as shown in **Figure S3.4, S3.5**) is challenging by gross visual inspection alone but can be realized through application of multivariate analytics algorithms that harness the full spectral information[64,65].

Figure 3.4A and **3.4B** shows the results of hierarchical cluster analysis (HCA)[45,66,67] for the shorter and the longer wavelength resonance peaks of SNPG, respectively. The dendrograms demonstrate that the shorter wavelength peak for SNPG around 440 nm is

extremely sensitive to the surrounding conditions and can be used to clearly distinguish between various dielectric media, notably water and 0.1 mmol/L BSA, which vary only by 0.001 in their RI. On the other hand, the peak around 580 nm is markedly less sensitive to the surrounding dielectric media and has difficulty in distinguishing between any of the aqueous solutions considered in our study. We attribute this to the differential near-field intensities experienced by the surrounding dielectric. Specifically, the surrounding media perceives a more pronounced near-field intensity for the silver nanoparticles, which is directly exposed to the media itself, as opposed to the gold nanoparticles that are covered by a 10 nm silica layer. A direct consequence of the differential sensitivities of the two peaks is that it enables the SNPG sample to be suitable for self-referenced measurements. The experimental finding is consistent with the results of our simulation study (further details are provided in the Supplementary Information, **Figure S3.6** & **S3.7**). **Figure 3.4C** plots the dendrograms showing the performance of STG samples in classifying the various concentrations. The STG sample, which has a single dominant resonant peak around 564 nm, is able to differentiate between the various concentrations although the accuracy is less than the SNPG sample. Thus, the SNPG sample is better suited for sensitive and self-referenced measurement of RI.
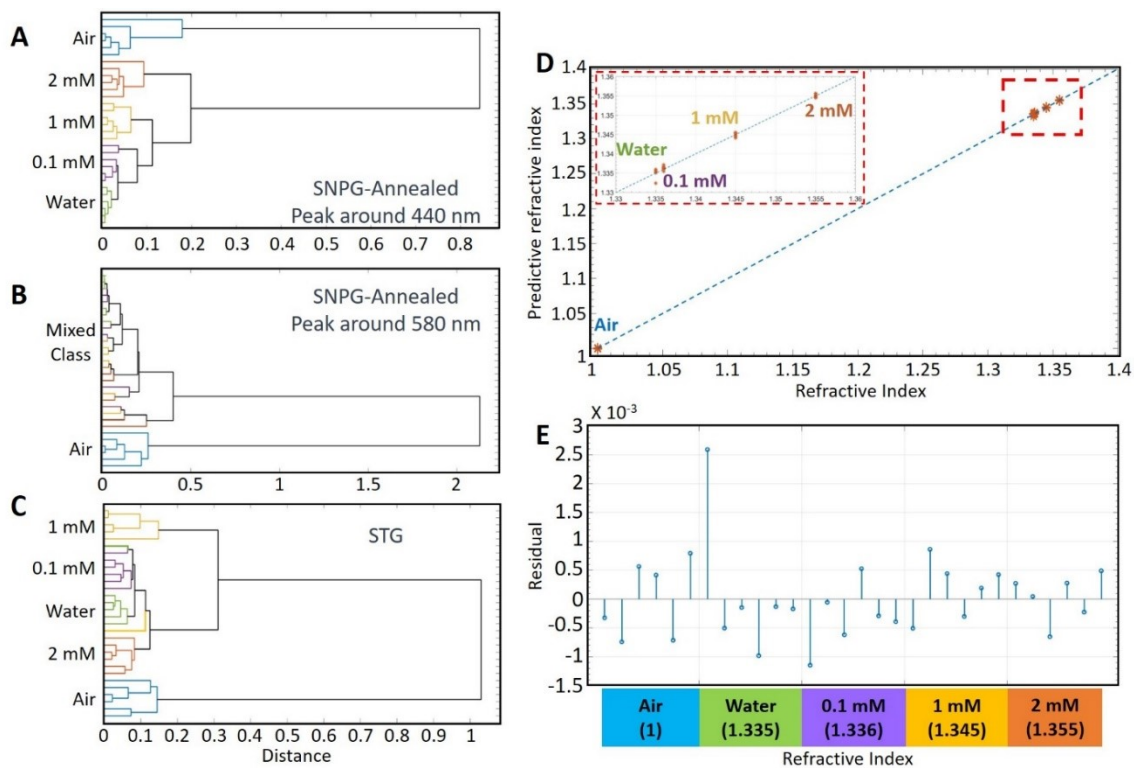
Figure 3.4 (A) Hierarchical cluster analysis of absorption spectra between 320-500 nm (surrounding the shorter wavelength resonance peak of SNPG) at various concentrations of BSA, air and water. The dendrogram indicates that the different solutions can be accurately identified. (B) Hierarchical cluster analysis of absorption spectra of longer wavelength resonance peak (500-800 nm) of SNPG at various concentrations of BSA, air and water. None of the solutions could be suitably classified. (C) Hierarchical cluster analysis of absorption spectra of resonance peak of STG at various concentrations of BSA, air and water. Different solutions could be identified although the accuracy is less compared to that obtained with the SNPG sample. (D) PLS prediction results of BSA solutions, water and air in the wavelength range of 320-500 nm (around the 440 nm peak region) in case of the SNPG-annealed substrate. The solid line denotes y=x and the red dashed box is an enlarged version of a part of the graph as shown in the subfigure. (E) PLS prediction residuals for absorption spectra of SNPG-annealed sample (around the 440 nm peak) belonging to the different RI.

Figure 3.4D shows the results of partial least squares (PLS) leave-one-out prediction

for the absorption spectra based on analysis of the SNPG-annealed 440 nm peak region.

The reference and PLS-predicted RI are given along the X and Y-axis, respectively. The

dashed line is plotted to illustrate y=x to explicitly understand the linearity of the response

(or the lack thereof). From the figure, it is evident that the predicted values show excellent

agreement with the reference concentrations and the corresponding root-mean-square error (RMSE) is calculated to be 0.0007. **Figure 3.4E** shows the prediction residual for spectra from various samples. It indicates that the SNPG substrate was able to accurately determine the RI down to the third decimal level, thereby underscoring its utility in monitoring subtle changes in analyte concentration in the surrounding milieu. To evaluate the linearity range of our measurement, we calculate the R-square value (coefficient of determination) between the predicted concentration and actual concentration, which is *ca.* 0.99. From **Figure S3.3** in the Supplementary Information, it can be concluded that within the range of our measurement, i.e. 0-2 mmol/L, the predicted BSA concentration shows a linear relationship with the refractive index. Using the calibration plot method discussed in the Materials & Methods, the LOD for BSA was computed to be 0.013 mmol/L.

CSP-based biomolecular detection with surface-enhanced Raman spectroscopy. The ability to design and fabricate nanostructures with a high degree of precision has transformed SERS from an esoteric method to a potent analytical tool over the last decade[38]. The surface properties directly impact the electromagnetic (and chemical) enhancement experienced by the molecules adsorbed on or in the vicinity of the nanostructured surface. Here, through proof-of-concept measurements in BSA solutions, we explored the utility of the CSP structures in recording SERS spectra of biomolecules.

We chose water as a representative solvent to mimic a normal biofluid measurement situation. In our investigation, we pipetted 10 μl of 10 μmol/L BSA solution on the nanostructures fabricated on silicon wafers and allowed it to dry under normal ambient condition. These silicon substrates have the same nanostructure arrangement as the glass substrates used in our RI sensing experiments but contribute lower fluorescence

backgrounds to the recorded SERS spectra. SERS signals were collected from the dried spot by focusing on the substrate surface. As a control experiment, Raman spectra were collected with the same concentration on a silicon wafer without any nanostructure. **Figure 3.5A** shows the spectra recorded from various kinds of substrates after baseline removal. The Raman spectra were normalized with the silicon background peak at 520 cm^{-1} [68]. Several vibrational peaks in the range of 900-1800 cm^{-1} are identified in the spectrum, 1001 cm^{-1} (phenylalanine), 1321 cm^{-1} (amide III), 1335 cm^{-1} (CH$_3$CH$_2$ wagging), 1446 cm^{-1} (CH$_2$ deformation), 1583 cm^{-1} (C=C bending mode of phenylalanine) and 1652 cm^{-1} (amide I)[68]. The experimental enhancement factor, as defined elsewhere in the literature[69] at 1335 cm^{-1} (CH$_3$CH$_2$ wagging) for the SNPG-annealed sample is about 22 whereas for STG is about 14. Thus, both the CSPs provide enhancement to the weak Raman signal of BSA, with the SNPG sample expectedly giving better enhancement. The enhancement factor obtained here is competitive with those obtained in the literature for similar measurement protocols[70] and, while substantially lower than those recorded from unstructured colloids, is more reproducible and has a well-defined spatial distribution. The latter is a key driver for future translation of such SERS platforms to the clinic or field setting. Furthermore, by comparing the intensities between signals from BSA solution on SNPG-annealed substrate and on (blank) silicon substrate, we determined the LOD value for our SERS measurements to be 0.096 μM. The linearity range of SERS study extends up to 0.5 mmol/L, i.e. the highest measured concentration in our pilot study. Our experimental range of measurements could be readily expanded; however, because our aim was to screen biomarkers in body fluids for diagnosis, higher concentrations are not deemed as biologically significant. To evaluate the reproducibility of the SERS measurements, the

coefficient of variation/relative standard deviation (RSD) was calculated at few of the major peaks in the Raman spectra (table provided in the Supplementary Information) and ranged between 11 to 19%.
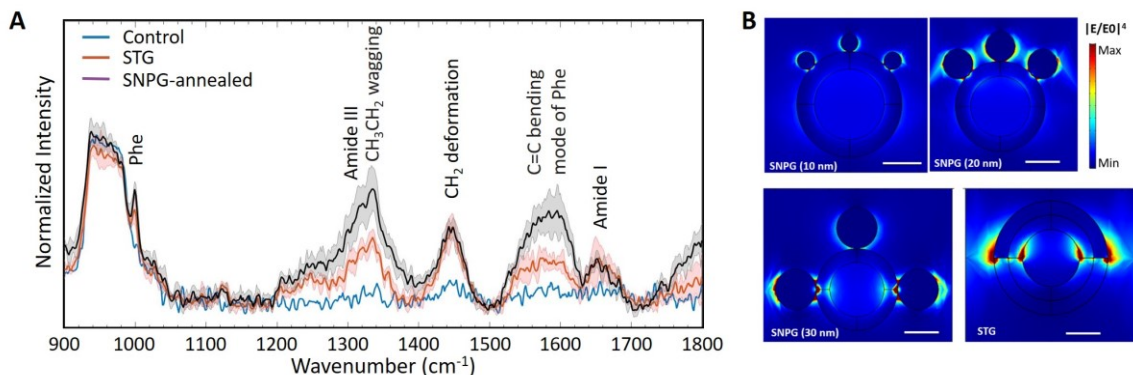
Figure 3.5 **SERS measurements of BSA solution and numerical calculation of field enhancement:** (A) SERS spectra of 10 μmol/L BSA on different substrates measured with a 2 second integration with a 530 nm laser in a micro-Raman system. Prominent Raman peaks are marked in the figure. The solid lines depict the average spectrum, and the shaded region represents the ± standard deviations (SD). (B) Distribution of $|E/E_0|^4$ in the XZ plane at 530 nm excitation. All the panels are plotted with the same range of color scale for comparison. The scale bar is 20 nm. The STG configuration shows significant enhancement at the sharp corners, which is due to the lightning rod effect but is absent in our experimental prototype.

In order to determine the origin of the SERS enhancement, using our previous model as in **Figure 3.2A**, we calculated the distribution of near field enhancement ($|E/E_0|^4$) in the XZ plane considering the analyte molecule to be most exposed to the enhanced field in this plane. A fourth power of electric field was chosen since SERS signal was proportional to the square of the intensity. **Figure 3.5B** plots the spatial variation of $|E/E_0|^4$ in the XZ plane for a wavelength of 530 nm, which is closer to the Raman excitation wavelength used in our experiments (532 nm). It can be clearly seen that the field enhancement covers a greater area for the SNPG (30 nm) arrangement, which allows for substantial interaction of analyte with the electromagnetic field leading to stronger SERS signals. The STG configuration

shows some enhancement at the corner of silver thin film due to the well-known lightning rod effect[11,69]. But, such sharp corners are absent in our experimental substrate; thus, enhancement in our experimental STG sample is lower than predicted. With near-IR laser excitation, it is expected that the gold nanoparticle will have more contribution relative to its silver counterpart in enhancing the SERS signal.

3.5 Conclusion

Localized surface plasmon resonance (LSPR) and surface-enhanced Raman spectroscopy (SERS) measurements offer important methods for label-free fingerprinting and quantitative biomolecular determination. The current study reports a large-area, versatile and self-referenced sensing platform, using a combination of metal and dielectric nanostructures, to perform composite scattering (i.e., elastic and inelastic scattering) measurements. Our design of CSP does not necessitate the use of lithography and is suitable for wafer scale fabrication. Using the CSP nanostructures, we show that LSPR sensing when combined with SERS offers dual-modality sensing that adds redundancy and encoding features, thus increasing measurement robustness and predictability. As a proof of concept, the SNPG design shows the potential as a means to self-reference. Notably, the inclusion of both the reference and sensing regions within the same physical space allows for a more compact design, and permits the use of a single detector rather than multiple detectors[32]. Crucially, the CSP does not require any physical or chemical alteration for its use as both a SERS probe and a LSPR sensor thereby underscoring its generalizability. In particular, our findings showcase the promise of the CSP as a sensing tool based on its demonstrated detection limit and the fact that Raman spectroscopy yields molecular-

specific information in the native state not readily obtainable from many other detection methods. While the preliminary results are promising, we expect additional improvement in detection limit and photonic throughput with further optimization of the probe design and corresponding sample fabrication procedures. Ultimately, we envision that the molecular specificity and multiplexing capability of the CSP will be leveraged for translating molecular markers into serum assays for accurate disease screening.

3.6 Appendix

1. Tunability of plasmon resonance peak (SNPG)
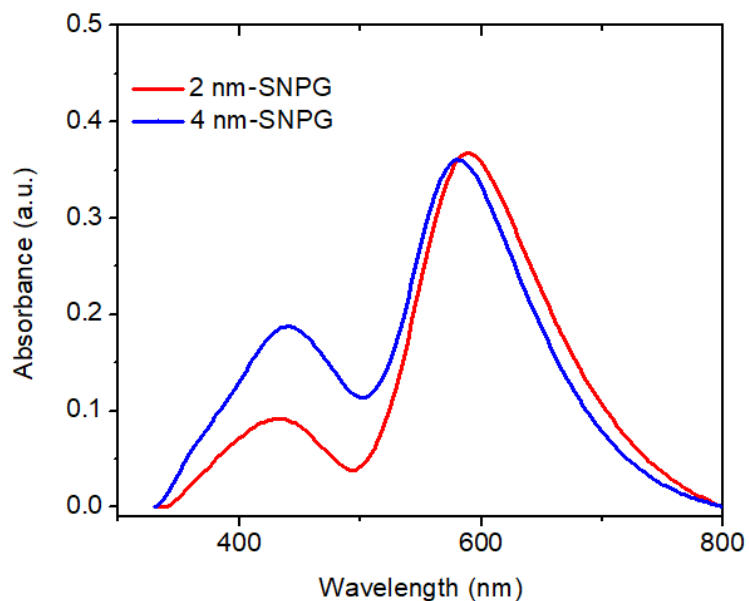
Figure S3.1 Experimental absorption spectra of SNPG sample with different thickness of silver layer (before annealing). Both the resonance peaks shift with change in thickness.
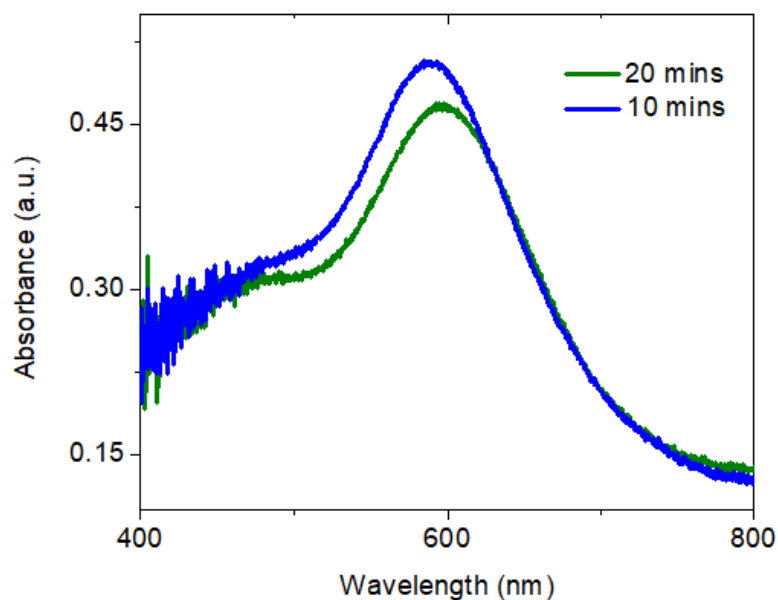
Figure S3.2 Experimental absorption spectra of 4 nm-SNPG sample with different annealing time. The longer wavelength resonance peak red shifts with increase in annealing time.

2. Refractive index measurement

Refractive index of water, 0.1 mM, 1 mM and 2 mM aqueous solution of BSA at 4^0C is measured using a scientific grade portable refractometer (Model: VEE GEE) under white light illumination. The refractometer has accuracy of ±0.0005.
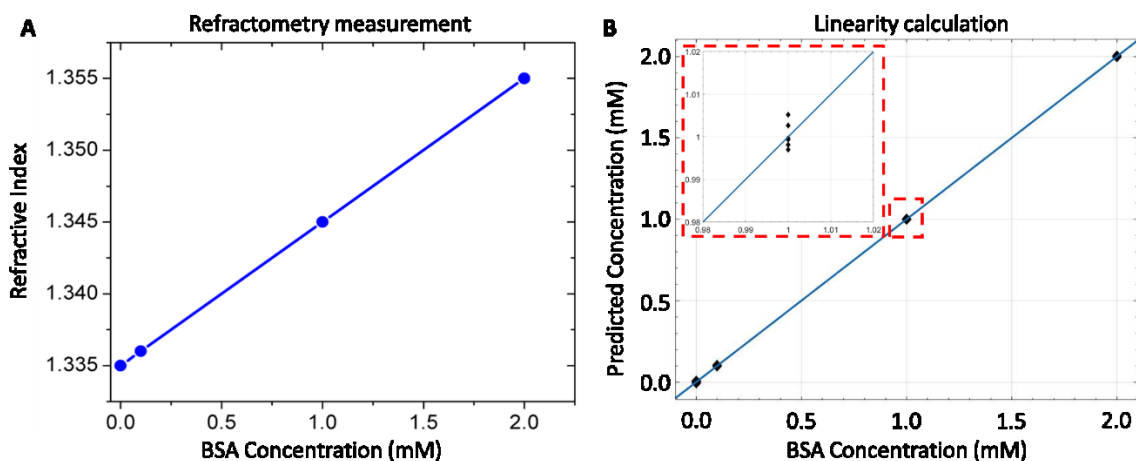
Figure S3.3 (A) Measured refractive index vs concentration of BSA. (B) PLS predicted concentration results of water and BSA solutions in the wavelength range of 320-500 nm (around the 440 nm peak region) in case of the SNPG-annealed substrate. The solid line denotes y=x. We have also provided an inset that zooms in on measurements at 1.0 mM BSA solution to clearly bring out the reproducibility, i.e. differences (or lack thereof) between repeated measurements.

3. Absorption spectra with different concentrations of BSA

72
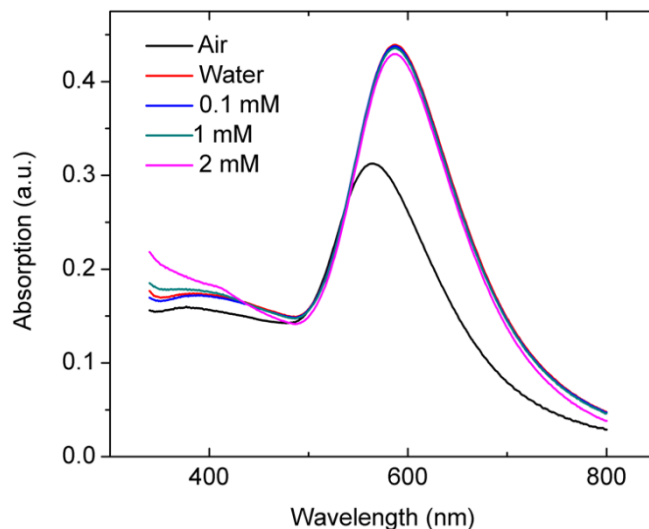
Figure S3.4 Absorption spectra of STG sample at different BSA concentrations.
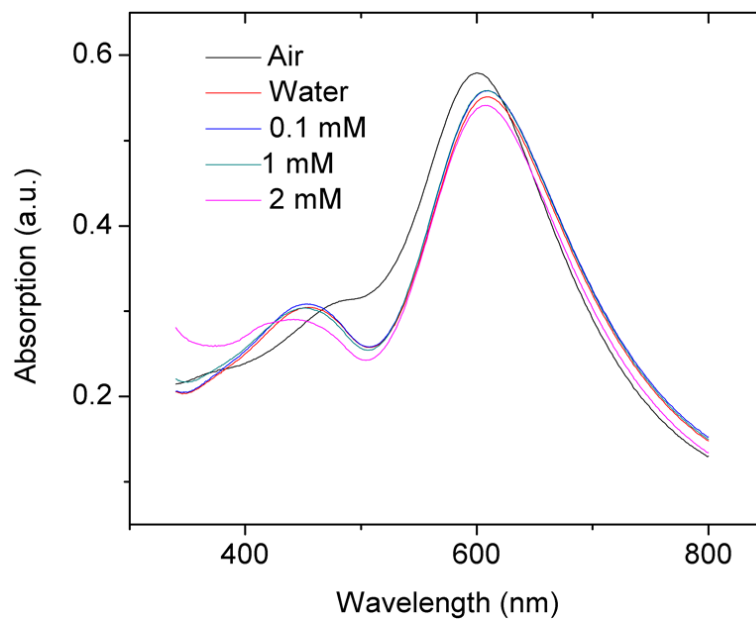
Figure S3.5 Absorption spectra of SNPG sample at different BSA concentrations.

4. Simulations
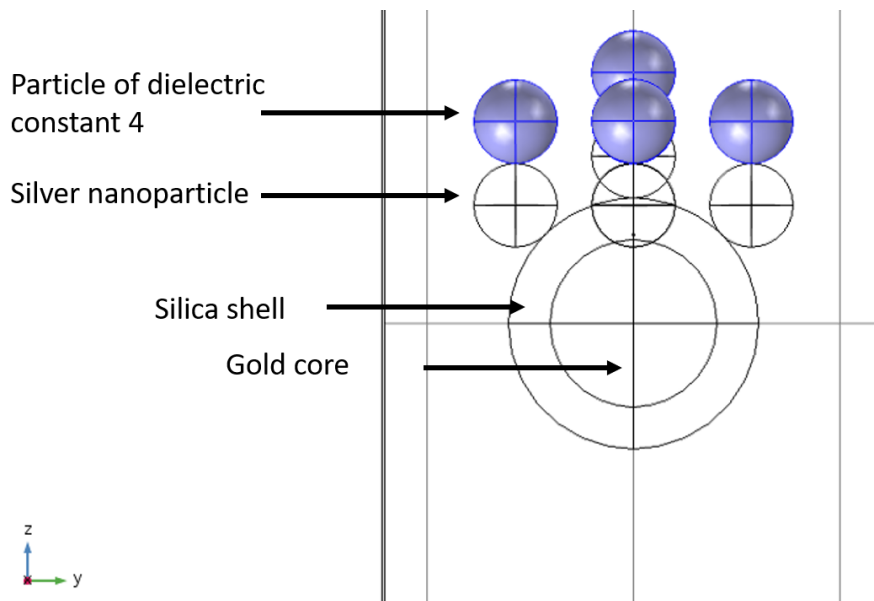
Figure S3.6 FEM simulation model with a non-metallic particle of dielectric constant 4 (represented in blue) considered in the vicinity of the silver nanoparticle.
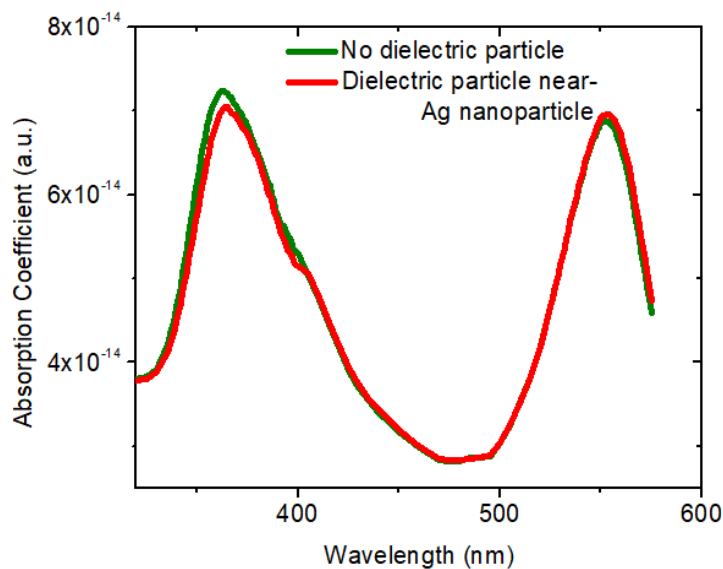
Figure S3.7 Absorption spectra demonstrating a shift in the silver nanoparticle resonance peak due to the presence of a dielectric particle whereas the peak due to gold nanoparticle remains unaffected. Substrate with multi plasmon resonance peaks such as this can be used for self-referenced measurement techniques.

5. SERS measurement
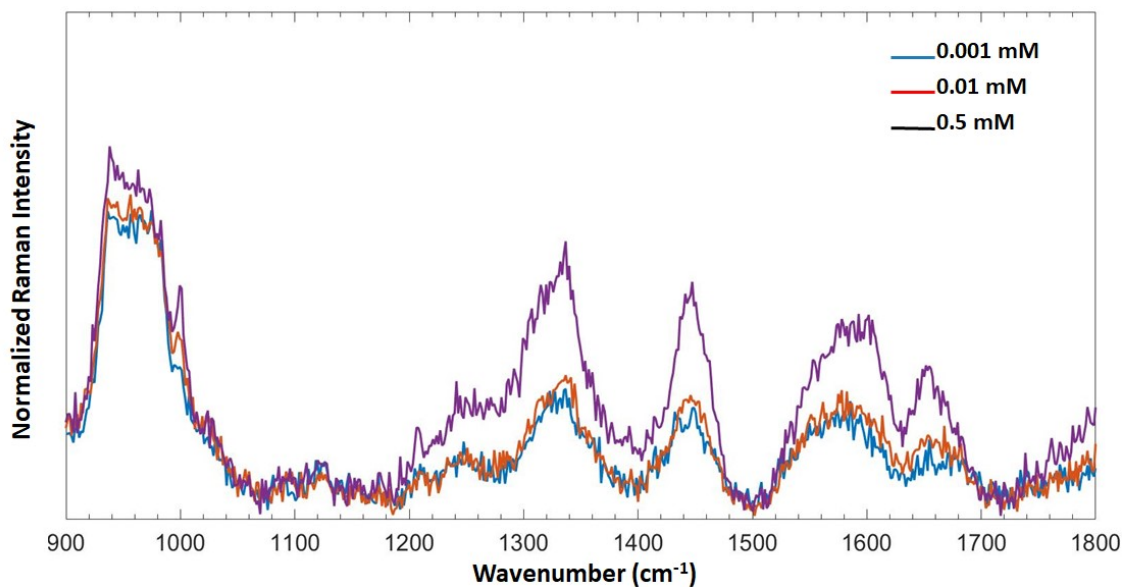
Figure S3.8 Raman spectra for different concentrations of BSA on SNPG sample. Strong Raman signal could be achieved for a very dilute concentration of BSA.
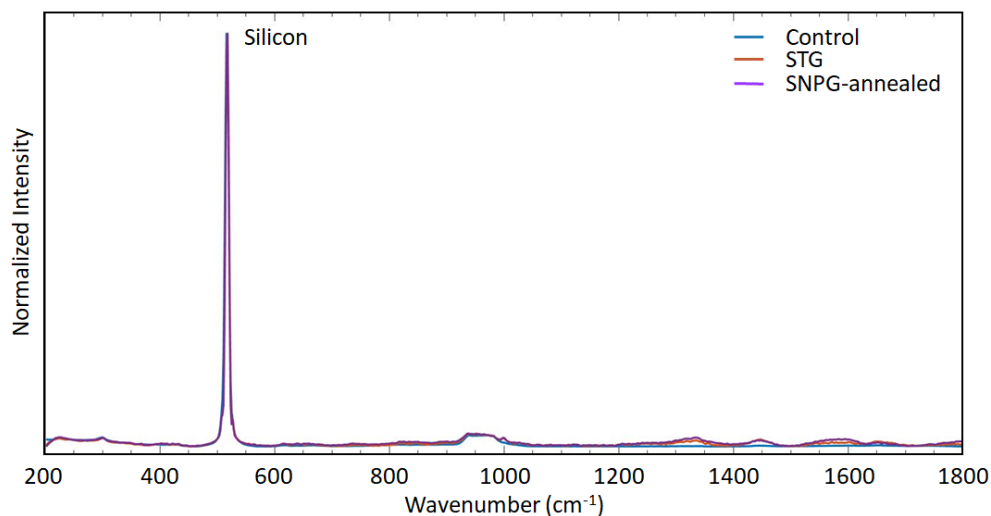
Figure S3.9 Raman spectra of 0.01 mmol/L BSA on different substrates with the Si peak at 520 cm^{-1} used for normalization.

6. SERS reproducibility

The spectra were collected from multiple positions across the area of the whole substrate to determine the reproducibility of our SERS measurements. The coefficient of variation at several prominent vibrational peaks of SNPG-annealed and STG probes are presented in the table below.

	Phe (1001 cm^{-1})	Amide III (1321 cm^{-1})	CH$_3$CH$_2$ wagging (1335 cm^{-1})
SNPG-annealed	0.11	0.18	0.19
STG	0.15	0.15	0.12

Acknowledgements

Bibliography

1. Mehrotra, P. Biosensors and their applications - A review. *J. Oral Biol. Craniofac. Res.*

2016, 6, 153–159.

2. Holzinger, M., Le Goff, A. and Cosnier, S. Nanomaterials for biosensing applications: a review. *Front Chem.* **2014**, 2, 63.

3. Bruchez, M., Jr. Semiconductor Nanocrystals as Fluorescent Biological Labels. *Science, 1998*, 2013–2016.

4. Zhao, X. *et al.* A rapid bioassay for single bacterial cell quantitation using bioconjugated nanoparticles. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 15027–15032.

5. Ferreira, H.A., Graham, D.L., Freitas, P.P. and Cabral, J.M.S. Biodetection using magnetically labeled biomolecules and arrays of spin valve sensors (invited). *J. Appl. Phys.* **2003**, 93, 7281–7286.

6. Sepúlveda, B., Angelomé, P.C., Lechuga, L.M. and Liz-Marzán, L.M. LSPR-based nanobiosensors. *Nano Today* **2009**, 4, 244–251.

7. Cheng, Y., Wang, M., Borghs, G. and Chen, H. Gold Nanoparticle Dimers for Plasmon Sensing. *Langmuir* **2011**, 27, 7884–7891.

8. Hao, F. *et al.* Symmetry Breaking in Plasmonic Nanocavities: Subradiant LSPR Sensing and a Tunable Fano Resonance. *Nano Lett.* **2008**, 8, 3983–3988.

9. Fan, X. *et al.* Sensitive optical biosensors for unlabeled targets: a review. *Anal. Chim. Acta* **2008**, 620, 8–26.

10. Maier, S.A. *Plasmonics: Fundamentals and Applications*. **2007**.

11. Schuller, J.A. *et al.* Plasmonics for extreme light concentration and manipulation. *Nat. Mater.* **2010**, 9, 193–204.

12. Liedberg, B., Nylander, C. and Lunström, I. Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators* **1983**, 4, 299–304.

13. Brolo, A. G. Plasmonics for future biosensors. *Nat. Photonics* **2012**, 6, 709–713.

14. Schwarz, B. *et al.* Monolithically integrated mid-infrared lab-on-a-chip using plasmonics and quantum cascade structures. *Nat. Commun.* **2014**, 5, 4085.

15. SadAbadi, H., Badilescu, S., Packirisamy, M. and Wüthrich, R. Integration of gold nanoparticles in PDMS microfluidics for lab-on-a-chip plasmonic biosensing of growth hormones. *Biosensors and Bioelectronics* **2013**, 44, 77–84.

16. Lu, X., Rycenga, M., Skrabalak, S.E., Wiley, B. and Xia, Y. Chemical Synthesis of Novel Plasmonic Nanoparticles. *Annu. Rev. Phys. Chem.* **2009**, 60, 167–192.

17. Stewart, M.E. *et al.* Nanostructured Plasmonic Sensors. *Chem. Rev.* **2008**, 108, 494–521.

18. Shen, Y. *et al.* Plasmonic gold mushroom arrays with refractive index sensing figures of merit approaching the theoretical limit. *Nat. Commun.* **2013**, 4, 2381 (2013).

19. Rindzevicius, T., Alaverdyan, Y., Käll, M., Andrew Murray, W. and Barnes, W. L. Long-Range Refractive Index Sensing Using Plasmonic Nanostructures. *J. Phys. Chem. C.* **2007**, 111, 11806–11810.

20. Gu, Z.Z. *et al.* Fabrication of a metal-coated three-dimensionally ordered macroporous film and its application as a refractive index sensor. *Angew. Chem. Int. Ed Engl.* **2002**, 41, 1154–1156.

21. Underwood, S. and Mulvaney, P. Effect of the Solution Refractive Index on the Color of Gold Colloids. *Langmuir* **1994**, 10, 3427–3430.

22. Mock, J.J., Smith, D.R. and Schultz, S. Local Refractive Index Dependence of Plasmon Resonance Spectra from Individual Nanoparticles. *Nano Lett.* **2003**, 3, 485–491.

23. McFarland, A.D. and Van Duyne, R.P. Single Silver Nanoparticles as Real-Time

Optical Sensors with Zeptomole Sensitivity. *Nano Lett.* **2003**, 3, 1057–1062.

24. Nehl, C.L., Liao, H. and Hafner, J.H. Optical properties of star-shaped gold nanoparticles. *Nano Lett.* **2006**, 6, 683–688.

25. Li, M., Kang, J.W., Dasari, R.R. and Barman, I. Shedding light on the extinction-enhancement duality in gold nanostar-enhanced Raman spectroscopy. *Angew. Chem. Int. Ed Engl.* **2014**, 53, 14115–14119.

26. Sherry, L.J. *et al.* Localized surface plasmon resonance spectroscopy of single silver nanocubes. *Nano Lett.* **2005**, 5, 2034–2038.

27. Burgin, J., Liu, M. and Guyot-Sionnest, P. Dielectric Sensing with Deposited Gold Bipyramids. *J. Phys. Chem. C* **2008**, 112, 19279–19282.

28. Mayer, K.M. *et al.* A label-free immunoassay based upon localized surface plasmon resonance of gold nanorods. *ACS Nano* **2008**, 2, 687–692.

29. Wang, H., Brandl, D.W., Le, F., Nordlander, P. and Halas, N.J. Nanorice: a hybrid plasmonic nanostructure. *Nano Lett.* **2006**, 6, 827–832.

30. Yong, Z., Lei, D.Y., Lam, C.H. and Wang, Y. Ultrahigh refractive index sensing performance of plasmonic quadrupole resonances in gold nanoparticles. *Nanoscale Res. Lett.* **2014**, 9, 187.

31. Sperling, J.R., Macias, G., Neale, S.L. and Clark, A.W. Multilayered Nanoplasmonic Arrays for Self-Referenced Biosensing. *ACS Appl. Mater. Inter.* **2018**, 10, 34774-34780.

32. Wersäll, M., Verre, R., Svedendahl, M., Johansson, P., Käll, M. and Shegai, T. Directional nanoplasmonic antennas for self-referenced refractometric molecular analysis. *J. Phys. Chem. C* **2014**, 118, 21075-21080.

33. Aslan, K., Lakowicz, J.R. and Geddes, C.D. Nanogold plasmon resonance-based glucose sensing. 2. Wavelength-ratiometric resonance light scattering. *Anal. Chem.* **2005**, 77, 2007–2014.

34. Monzón-Hernández, D., Villatoro, J., Talavera, D. and Luna-Moreno, D. Optical-fiber surface-plasmon resonance sensor with multiple resonance peaks. *Appl. Opt.* **2004**, 43, 1216–1220.

35. Lin, V.K. *et al.* Dual wavelength sensing based on interacting gold nanodisk trimers. *Nanotechnology* **2010**, 21, 305501.

36. Dondapati, S.K. *et al.* Label-free biosensing based on single gold nanostars as plasmonic transducers. *ACS Nano* **2010**, 4, 6318–6322.

37. Nie, S. Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering. *Science* **1997**, 275, 1102–1106.

38. Stiles, P.L., Dieringer, J.A., Shah, N.C. and Van Duyne, R.P. Surface-enhanced Raman spectroscopy. *Annu. Rev. Anal. Chem.* **2008**, 1, 601–626.

39. Li, M., Cushing, S.K. and Wu, N. Plasmon-enhanced optical sensors: a review. *Analyst* **2015**, 140, 386–406.

40. Tokel, O., Inci, F. and Demirci, U. Advances in plasmonic technologies for point of care applications. *Chem. Rev.* **2014**, 114, 5728–5752.

41. Im, H., Bantz, K.C., Lee, S.H., Johnson, T.W., Haynes, C.L. and Oh, S.H. Self-assembled plasmonic nanoring cavity arrays for SERS and LSPR biosensing. *Adv. Mater.* **2013**, 25, 2678-2685.

42. Potara, M., Gabudean, A.M. and Astilean, S. Solution-phase, dual LSPR-SERS plasmonic sensors of high sensitivity and stability based on chitosan-coated anisotropic

silver nanoparticles. *J. Mater. Chem.* **2011**, 21, 3625-3633.

43. Focsan, M., Craciun, A.M., Potara, M., Leordean, C., Vulpoi, A., Maniu, D. and Astilean, S. Flexible and tunable 3D gold nanocups platform as plasmonic biosensor for specific dual LSPR-SERS immuno-detection. *Sci. Rep.* **2017**, 7, 14240.

44. Byrne, B., Stack, E., Gilmartin, N. and O'Kennedy, R. Antibody-based sensors: principles, problems and potential for detection of pathogens and associated toxins. *Sensors* **2009**, 9, 4407–4445.

45. Johnson, S.C. Hierarchical clustering schemes. *Psychometrika* **1967**, 32, 241–254.

46. Haaland, D.M. and Thomas, E.V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, 60, 1193-1202.

47. Schafer, R. What Is a Savitzky-Golay Filter? *IEEE Signal Process. Mag.* **2011**, 28, 111-117.

48. Currie, L.A. International Recommendations Offered on Analytical Detection and Quantification Concepts and Nomenclature: Preamble, in Validation of Analytical Methods. *Anal. Chim. Acta* **1999**, 391, 103-134.

49. Yang, S., Dai, X., Stogin, B.B. and Wong, T.S. Ultrasensitive surface-enhanced Raman scattering detection in common fluids. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, 113, 268-273.

50. Palik, E.D. *Handbook of Optical Constants of Solids*. (Academic Press, **2012**).

51. Tesler, A.B., Chuntonov, L., Karakouz, T., Bendikov, T.A., Haran, G., Vaskevich, A. and Rubinstein, I. Tunable localized plasmon transducers prepared by thermal dewetting of percolated evaporated gold films. *J. Phys. Chem. C* **2011**, 115, 24642-

24652.

52. Hawkeye, M.M. and Brett, M.J. Glancing angle deposition: fabrication, properties, and applications of micro-and nanostructured thin films. *J. Vac. Sci. Technol. A* **2007**, 25, 1317-1335.

53. Gish, D.A., Nsiah, F., McDermott, M.T. and Brett, M.J. Localized surface plasmon resonance biosensor using silver nanostructures fabricated by glancing angle deposition. *Anal. Chem.* **2007**, 79, 4228–4232.

54. Zhao, Y.P., Ye, D.X., Wang, G.C. and Lu, T.M. Novel Nano-Column and Nano-Flower Arrays by Glancing Angle Deposition. *Nano Lett.* **2002**, 2, 351–354.

55. Singh, J.H., Nair, G., Ghosh, A. and Ghosh, A. Wafer scale fabrication of porous three-dimensional plasmonic metamaterials for the visible region: chiral and beyond. *Nanoscale* **2013**, 5, 7224–7228.

56. Paria, D. *et al.* Ultrahigh field enhancement and photoresponse in atomically separated arrays of plasmonic dimers. *Adv. Mater.* **2015**, 27, 1751–1758.

57. Romero, I., Aizpurua, J., Bryant, G.W. and García De Abajo, F.J. Plasmons in nearly touching metallic nanoparticles: singular response in the limit of touching dimers. *Opt. Express* **2006**, 14, 9988–9999.

58. Brongersma, M.L., Halas, N.J. and Nordlander, P. Plasmon-induced hot carrier science and technology. *Nat. Nanotechnol.* **2015**, 10, 25–34.

59. Liu, X. *et al.* Tunable Dipole Surface Plasmon Resonances of Silver Nanoparticles by Cladding Dielectric Layers. *Sci. Rep.* **2015**, 5, 12555.

60. Zhang, X., Zhang, J., Wang, H., Hao, Y., Zhang, X., Wang, T., Wang, Y., Zhao, R., Zhang, H. and Yang, B. Thermal-induced surface plasmon band shift of gold

nanoparticle monolayer: morphology and refractive index sensitivity. **2010**, *Nanotechnology* 21, 465702.

61. Szunerits, S. and Boukherroub, R. Sensing using localised surface plasmon resonance sensors. *Chem. Commun.* **2012**, 48, 8999–9010.

62. Sagle, L.B., Ruvuna, L.K., Ruemmele, J.A. and Van Duyne, R.P. Advances in localized surface plasmon resonance spectroscopy biosensing. **2011**, *Nanomedicine* 6, 1447-1462.

63. Muri, H., Bano, A. and Hjelme, D. LSPR and interferometric sensor modalities combined using a double-clad optical fiber. **2018**, *Sensors* 18, 187.

64. Naeser, K. and Hjortdal, J. Multivariate analysis of refractive data: mathematics and statistics of spherocylinders. *J. Cataract Refract. Surg.* **2001**, 27, 129–142.

65. Godinho, R.B., Santos, M.C. and Poppi, R. J. Quality control of fragrances using Raman spectroscopy and multivariate analysis. *J. Raman Spectrosc.* **2015**, 47, 579–584.

66. Sener, G., Uzun, L. and Denizli, A. Colorimetric Sensor Array Based on Gold Nanoparticles and Amino Acids for Identification of Toxic Metal Ions in Water. *ACS Appl. Mater. Interfaces* **2014**, 6, 18395–18400.

67. Tomita, S., Ishihara, S. and Kurita, R. A Multi-Fluorescent DNA/Graphene Oxide Conjugate Sensor for Signature-Based Protein Discrimination. *Sensors* **2017**, 17, 2194.

68. Novara, C. *et al.* SERS-Active Ag Nanoparticles on Porous Silicon and PDMS Substrates: A Comparative Study of Uniformity and Raman Efficiency. *J. Phys. Chem. C* **2016**, 120, 16946–16953.

69. Liu, Y., Xu, S., Li, H., Jian, X. and Xu, W. Localized and propagating surface plasmon

co-enhanced Raman spectroscopy based on evanescent field excitation. *Chem. Commun.* **2011**, 47, 3784–3786.

70. Paria, D. *et al.* Graphene–silver hybrid devices for sensitive photodetection in the ultraviolet. *Nanoscale* **2018**, 10, 7685–7693.

Chapter 4. Label-free Raman spectroscopy provides early determination and precise localization of breast cancer-colonized bone alterations

4.1 Abstract

Breast neoplasms frequently colonize bone and induce development of osteolytic bone lesions by disrupting the homeostasis of the bone microenvironment. This degenerative process can lead to bone pain and pathological bone fracture, a major cause of cancer morbidity and diminished quality of life, which is exacerbated by our limited ability to monitor early metastatic disease in bone and assess fracture risk. Spurred by its label-free, real-time nature and its exquisite molecular specificity, we employed spontaneous Raman spectroscopy to assess and quantify early metastasis driven biochemical alterations to bone composition. As early as two weeks after intracardiac inoculations of MDA-MB-435 breast cancer cells in NOD-SCID mice, Raman spectroscopic measurements in the femur and spine revealed consistent changes in carbonate substitution, overall mineralization as well as crystallinity increase in tumor-bearing bones when compared with their normal counterparts. Our observations reveal the possibility of early stage detection of biochemical changes in the tumor-bearing bones – significantly before morphological variations are captured through radiographic diagnosis. This study paves the way for a better molecular understanding of altered bone remodeling in such metastatic niches, and for further clinical studies with the goal of establishing a non-invasive tool for early metastasis detection and prediction of pathological fracture risk in breast cancer.

4.2 Introduction

In the United States, breast cancer remains the most common malignant neoplasm of women and it is estimated that ~250,000 cases of female breast carcinoma *in situ* will be diagnosed along with an estimated 40,000 deaths in 2017[1]. While local breast tumors respond very well to therapy, distant metastases are a frequent complication with a very poor response to therapy[2]. The highest incidence of metastatic progression is to bone with the latter representing the first metastatic site in 30-40% of breast cancer patients[3,4], and bone metastases being reported in up to nearly 70% of patients during disease progression[4,5]. There are three types of bone metastases: osteolytic, marked by heightened resorption that results in loss of bone mineral and matrix; osteoblastic that causes an increase in abnormal weakened bone formation; and mixed osteolytic/osteoblastic lesions[6]. While osteolytic disease predominates[7], all three forms occur in the context of breast cancer[8]. Bone colonization and dysregulation of the normal bone remodeling processes results in a range of skeletal related events (SREs), such as severe bone pain, hypercalcemia, ablation of bone marrow resulting in pancytopenia, spinal cord compression as well as pathological fractures[5,6,8,9]. Such SREs often occur in load-bearing bones, for instance the neck of the femur or in the pelvis, which can present particular treatment challenges.

Given these circumstances, early accurate identification of patients at risk for bone metastasis is a critical need for the correct staging of patients and selection of appropriate therapeutic regimens[6,10]. Current imaging technologies used to screen for bone metastatic lesions include ^{99m}Tc based bone scintigraphy (BS), single photon emission computed tomography (SPECT), alone or combined with computed tomography (CT), CT combined with positron emission tomography (PET), and whole body magnetic resonance imaging

(wbMRI)[10]. Advances in PET/CT have provided spatial resolutions in the 5 to 8 mm (full-width at half-maximum (FWHM)) range[11], which makes detection of small early lesions more likely. Nevertheless, such detection does not provide information about the bone's microstructural composition and whether it has been compromised.

To provide comprehensive management of bone metastases requires not only a determination of metastatic lesions but also an accurate assessment of fracture risk[6]. Fracture risk needs to be assessed at the time of discovery as well as monitored during follow up of response to treatment, as a decrease in tumor burden alone does not ensure that a corresponding increase in bone integrity and strength has occurred[12]. Risk prediction based on bone mineral density (BMD) using dual energy X-ray absorptiometry (DXA) has been a standard of practice. However, this methodology does not provide an accurate estimate of bone mineral content[13]. Computed tomography (CT) is superior to 2D X-ray technologies for evaluating BMD and a recently developed CT-based structural rigidity analysis technology has demonstrated improved fracture risk prediction with two clinical trials showing 100% sensitivity but only fair, 60-70%, specificity[12,14]. Numerous recent reports have discussed why an evaluation of BMD alone falls short of accurately gauging bone fragility by pointing out that the organic component of bone, in conjunction with bio-hydroxyapatite, contributes greatly to bone toughness[15-18]. An accurate fracture risk assessment demands knowledge of the underlying biochemical matrix integrity along with the crystalline mineral composition of the bone. Hence, development of non-invasive technologies that can detect changes in bone matrix and mineral composition at early stages of colonization would meet a significant clinical need.

Label-free vibrational spectroscopy's unique attributes can address these unmet needs as it provides objective biomarkers of bone composition for diagnoses[16,18] and may provide patient stratifications for more effective therapy[17]. The exquisite molecular specificity of this approach enables multiplexed biomolecular analyses without necessitating exposure to radiation or addition of exogenous contrast agents[19,20]. Furthermore, in comparison to IR spectroscopy, the higher spatial resolution and the ability to analyze fresh tissue specimen opens up tremendous opportunities for Raman scattering measurements. A recent review indicates the wide-ranging potential of Raman spectroscopy to assess the biochemical attributes of bone, which when combined with mechanical loading regimes, correlate with bone failure responses at the ultrastructural level[21]. In addition, recent studies have expanded on this and demonstrated Raman spectroscopy's ability to detect specific pathological changes in bone matrix components, including alterations in phosphate, carbonate, the amide backbone of collagen, as well as collagen cross-link maturity, and use them in predictive models for fracture risk[22-24]. Important insights can also be gained from the emerging evidence on Raman spectral changes in metastatic bone primed by prostate and breast cancer[16,25,26].

Here, we present results from a pilot study designed to determine whether Raman spectroscopy can detect changes in mineral and biomolecular components of bone early in metastatic progression. The above-cited studies on the use of Raman spectroscopy to evaluate bone matrix changes in metastatic model systems performed the assessments at mid-to-late stage disease, and a time-related assessment of bone quality alteration, particularly at early colonization, remains unexplored. We used a well-established bone metastatic model system where intracardiac injections of MDA-MB-435 breast cancer cells

results in bone metastases[27-29]. During this study, bone metastasis progression was tracked weekly for five weeks. Raman spectral evaluations were performed on *ex vivo* specimens of femurs and spines (**Figure 4.1**). Relative to normal bone, Raman spectroscopy was able to detect changes in the biochemical characteristics of bones with metastatic lesions as early as two weeks after tumor cell inoculations while X-ray images showed no sign of disease even after 5 weeks of metastatic progression. Our findings demonstrate the feasibility of using Raman spectroscopy for early detection and localization of metastatic disease in bone by quantifying surrogate markers, notably changes in bone matrix composition, at the site of the disease. These early alterations in the intrinsic biochemical characteristics of bone suggest compromised integrity and a weakening of the bone. These findings pave the way for further investigations into pathological fracture risk estimation through noninvasive, spatially offset Raman spectroscopy measurements[30]) of bone in real time.
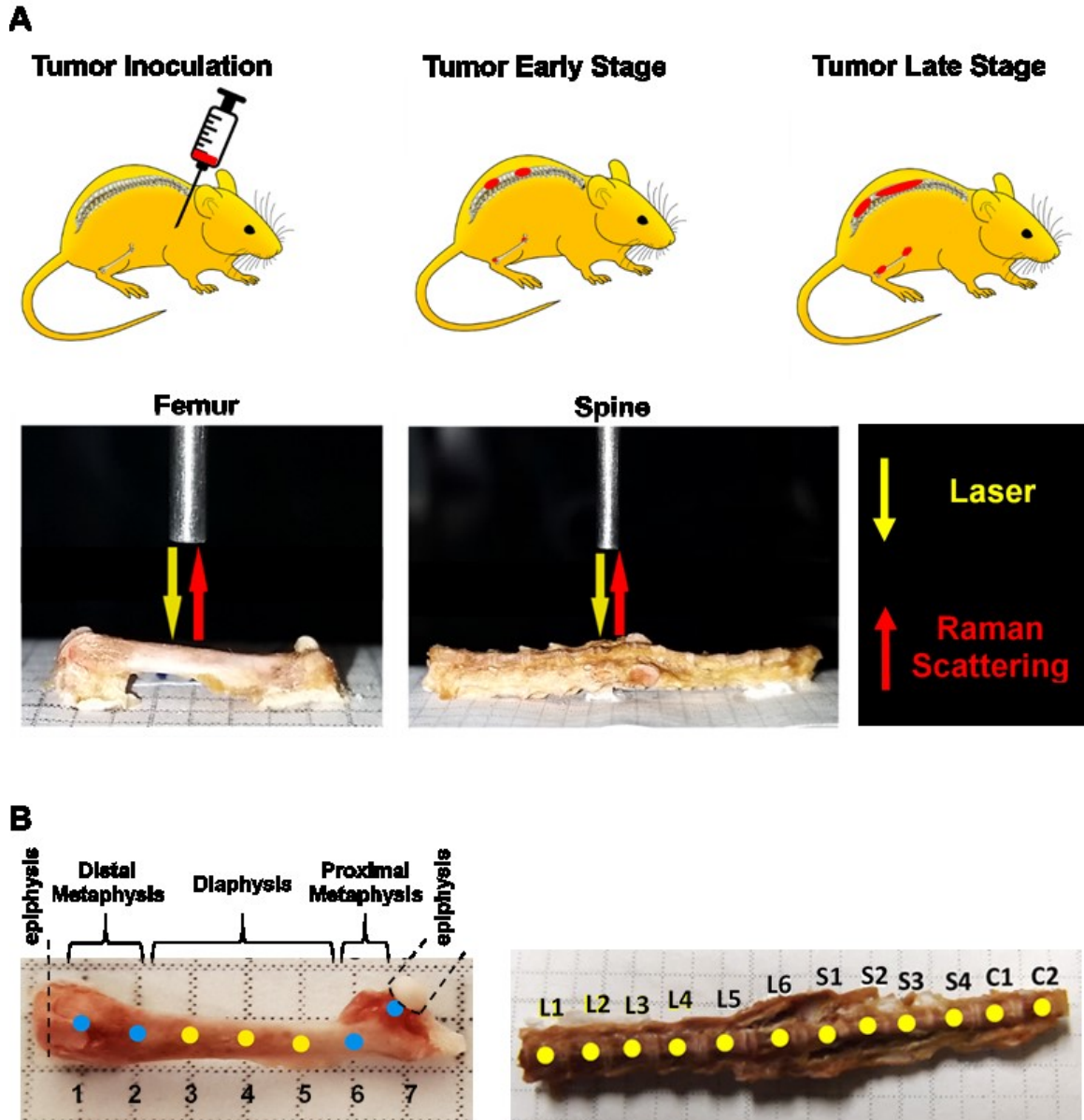
Figure 4.1 **Schematic representation of experimental model of breast cancer bone metastasis and depiction of subsequent Raman spectroscopy measurements.** (A) Intracardiac injected of breast cancer 435-tdT cells (top: left-hand panel) and ensuing metastases in the femur and spine as demarcated in red (top: central and right-hand panels). Raman microspectroscopy (bottom panels) was used to record spectra from these affected femurs and spines. (B) Raman spectra were collected at 2 mm intervals along the length of the femurs as indicated by numbered spots. Raman spectra of the spines were collected from central regions of lumbar (L1-L6), sacral (S1-S4), and caudal (C1-C2) vertebrae.

4.3 Experimental section

Cell culture. The human breast cancer cell line, MDA-MB-435, was obtained from ATCC. The MDA-MB-435 cell line was established in 1976 from a pleural effusion from an untreated 31-year-old female diagnosed with adenocarcinoma of the breast[31]. MDA-MB-435 cells were authenticated at the Johns Hopkins Genetic Resource Core Facility with the short tandem repeat marker results cross checked against cell lines at the ATCC bank. Generation and characterization of MDA-MB-435 engineered to constitutively express a bright red fluorescent protein: tdTomato (designated in the further discussion as 435-tdT cells), to facilitate *in vivo* and *ex vivo* tracking of metastatic progression, has been previously described[32]. Cells were cultured in RPMI-10% FBS medium in a standard humidified incubator at 37º C and 5% CO_2.

Intracardiac experimental bone metastasis model. 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 were used. Four six-to-eight week old mice were anesthetized by intraperitoneal injections of a mixture of xylazine (11 mg/kg) and ketamine (72 mg/kg). Mice were successfully injected with 435-tdT cells ($1x10^5$) in 0.1 ml of sterile RPMI-10% FBS through the intercostal space into the left ventricle of the heart using 26-gauge needle on a tuberculin syringe. Mice were sacrificed at different time points: on the day of tumor cell injection and at two, four and five weeks post tumor cell injection, by administering an overdose of anesthetic [saline:ketamine:acepromazine (2:1:1)] followed by cervical dislocation. Intact skeletons were then immediately dissected away

from soft tissue, wrapped in phosphate buffered saline soaked gauze and then in aluminium foil and stored at -20º C.

Live animal optical and X-ray imaging and *ex vivo* fluorescence microscopy. Live animal optical imaging was done in a Xenogen IVIS® Spectrum (PerkinElmer) optical scanner under 2% isoflurane/O_2 gas anesthesia, as previously described[32]. The spectral unmixing tool in the Living Image® 4.2 software package was used to remove background autofluorescence. The unmixing tool also provided a means to focus on fluorescence from bone by masking the fluorescence from brain metastases that were also generated with this mouse model. X-ray images were captured using a Faxitron® MX-20 X-ray scanner (Faxitron X-ray Corp., Lincolnshire, IL).

Fluorescence microscopy was on an inverted Nikon ECLIPSE TS 100 microscope (Nikon Instruments, Inc., Melliville, NY) equipped with a Texas Red filter cube. The fluorescence light source was an X-Cite 120 Fluorescence Illumination System (Photonic Solutions, Inc., Edinburgh, UK). A 2x objective and 2s exposure time were used. Anterior and posterior photomicrographs of each end of the femurs were obtained using a Photometrics CoolSnap™ ES (Roper Scientific, Trenton, NJ) camera interfaced with NIS-Elements F3.2 software. Fluorescence intensities of approximately equal sized ROIs of the anterior and posterior metaphysis regions, which excluded the non-fluorescing diaphysis, were quantitated with ImageJ. Images of proximal and distal metaphysis, which contained overlapping portions of each diaphysis, were stitched together during reconstruction.

Raman spectroscopy. Intact femurs and spines harvested from control mice without tumors and 435-tdT tumor-bearing mice were placed on a flat aluminum substrate. Baseline Raman spectra of intact vertebrae and femurs from control mice were obtained prior to

analyses of tumor burdened bones. Bones were stripped of the periosteum by lightly scraping with a scalpel and were wetted with Dulbecco's phosphate buffered saline. For the femurs, measurements from the distal metaphysis through the diaphysis to the proximal metaphysis were made at 2 mm intervals along the bone (**Figure 4.1A & 4.1B**) resulting in 35 spectra per femur. For the spines, measurements were taken at approximately the center of lumbar (L1-L6), sacral (S1-S4), and caudal (C1-C2) vertebrae (**Figure 4.1A & 4.1B**) resulting in 60 spectra per vertebral column. A home-built fiber-probe based Raman spectroscopy system was used to record the spectral profiles with 300s exposure time[33]. Briefly, an 830 nm diode laser (Process Instruments, Salt Lake City, UT) was used as the excitation source. A lensed fiber-optic Raman bundled contact probe (EmVision LLC, FL) was used to deliver the excitation beam through its central fiber (200 μm core) and acquire the back-scattered light through an annular ring of optical fibers (300 μm core). The scattered light was transmitted by the fiber-probe to a f/1.8i spectrograph (Holospec, Kaiser Optical Systems, MI) while a thermoelectrically cooled deep-depletion CCD camera (PIXIS 400BR, Princeton Instruments, NJ) was used for spectral recording. The laser power delivered to the sample surface was maintained at approximately 15 mW. Measurements over the length of the bone were enabled by scanning the fiber-probe using motorized translation stages (T-LS13M, Zaber Technologies, Inc., Vancouver, Canada). Zaber console (open-source software) was employed to control the raster scan as well as maintain a constant distance above the bone surface. Five spectra were collected from each measurement site.

Data analysis. The acquired spectra were imported into the MATLAB 2013a (Mathworks, Inc., Natick, MA) environment for further analysis. Spectra were processed

to remove interference from cosmic rays. A background spectrum, obtained from measurements on the aluminum substrate, was subtracted from the acquired spectra to correct for extraneous optical fiber-probe background signal. The resultant spectra were divided by the white light response signal (obtained from a $BaSO_4$ sample under tungsten halogen lamp illumination) to compensate for any spectral non-uniformity in the detection. Next, the fluorescence background was removed using an automated method outlined by the Berger laboratory[34]. The resultant spectra were normalized to the intensity of the PO_4^{3-} v1 peak (*ca.* 958 cm^{-1}).

In addition to peak height and full width at half maximum (FWHM) analyses, the Raman spectra were subjected to principal component analyses (PCA). Operating without any *a priori* knowledge of the samples, PCA seeks to project the spectral data onto a set of linearly uncorrelated (orthogonal) directions, *i.e.,* principal components (PC), such that the variance in the original data can be captured using only a few PCs[35]. Furthermore, support vector machines (SVM) were used on the PC score inputs to develop a decision algorithm for spectroscopically predicting the progression of tumor-induced changes in the bone. SVM is a supervised learning model that is built on structural risk minimization concepts and can efficiently perform non-linear classification by implicitly mapping the inputs into high-dimensional feature spaces through a kernel. A radial basis function (RBF) with a Gaussian envelope was chosen as the kernel, and the kernel parameters were optimized based on an automated grid search algorithm[36-38]. The output of the SVM-derived decision algorithm was validated against the known class labels, *i.e.*, the time point of the harvested bone sample. The performance of the algorithms was evaluated by determining the classification accuracy in a leave-one-spectrum-out cross-validation protocol.

4.4 Results and Discussion

In order to track metastatic progression, we employed a variant of MDA-MB-435 breast cancer cells that have been engineered to constitutively express a bright red fluorescent protein[32]. As shown in **Figure 4.2A**, fluorescence signals from bone, e.g., spine and scapulae, were recorded at weeks 4 and 5 and, at the latter time, fluorescence from the left pelvis/proximal femur region was also seen. No fluorescence signals were seen at week 2 (**Figure 4.2A**). Thus, optical imaging could not detect (as shown below and by a previous group[28]) the nascent metastatic disease already presents at week 2. In addition, as shown in **Figure 4.2B**, no metastatic lesions were revealed by X-ray imaging of the femurs at any time point (or in the spine specimen, data not shown). This is consistent with the low sensitivity of X-ray imaging in the clinic where bone metastases are reported to be only detected in those cases where 30-75% of skeletal destruction has already occurred[14].

This qualitative imaging data provided a means to establish the time course of metastatic progression and a confirmation that, similar to the clinical situation, X-ray imaging offers inadequate assessment of bone metastases, even when optical imaging indicated substantial tumor burden in bone (**Figure 4.2A**, week 4 & 5). As such, we hypothesized that this model system would provide a suitable sample set to test whether Raman spectroscopy can detect early alterations of bone matrix integrity, which may be considered as a surrogate assessment of the underlying metastatic involvement.
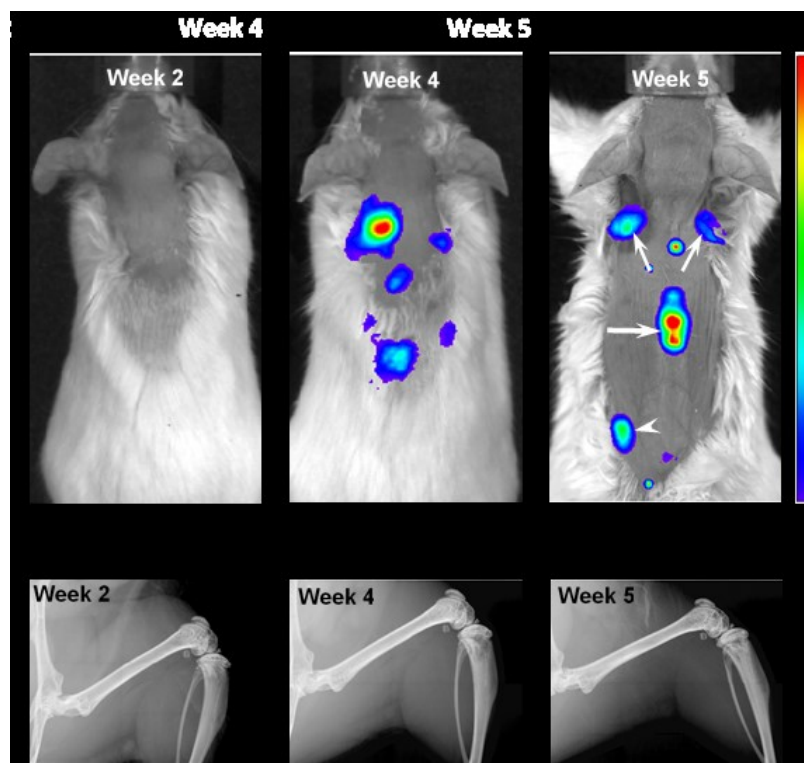
Figure 4.2 **Live animal optical and x-ray imaging.** (A) Fluorescence imaging of tdTomato signals from tumor-bearing mouse at week 2 (left panel) and tumor-bearing mice at 4 and 5 weeks post 435-tdT inoculations. In the week 5 image, fluorescent metastatic lesions in the scapulae (small arrows), lower thoracic-upper lumbar region of the spine (large arrow) and left proximal femur/pelvis region (arrowhead) were evident. Note, to better ascertain bone fluorescence signals, intense brain fluorescence was masked in the middle and right panel images. (B) X-ray images from a Faxitron x-ray scanner displaying the femur of the same mice shown in the corresponding panels in (A). No metastatic bone lesions were revealed in any of the x-ray images.

We focused our Raman spectroscopic analyses on femurs and vertebrae as the former is the most affected long bone and spine represents one-third of the total skeletal metastases observed in breast cancer patients[8,9]. Consequently, metastatic involvement at either or both sites accounts for a high proportion of SREs, and an early evaluation of metastases at these sites remains an unmet need. Assessing bone integrity using Raman spectroscopy provides specific quantitative evaluations of the bio-hydroxyapatite (mineral) component simultaneously with the associated collagen component (matrix)[21,23,25,26]. The mineral component is often calcium-deficient, and possesses carbonate substitutions with respect

to stoichiometric hydroxyapatite. Recent investigations have extended the assessments of bone fragility and metastatic lesions using Raman spectroscopy to include in-depth evaluations of collagen integrity[16,18,24] along with the contributions of proteoglycans, tissue water, and lipids[17] as well as oxidative damage[39].

During this pilot study, we focused on peak signals associated with bio-hydroxyapatite to collagen matrix content, mineral crystallinity, and carbonation to derive specific quantitative distinctions between control and tumor-bearing bones. **Figure 4.3** shows representative Raman spectra from femurs and spines from a non-tumor bearing normal control mouse (blue tracings) and a tumor-bearing mouse (red tracings) sacrificed 5 weeks after tumor cell inoculations. The distinctive Raman peaks at *ca.* 958, 1004, 1070, 1250 and 1450 cm^{-1} correspond to the vibrational modes of phosphate $\nu 1$, phenylalanine with potential contributions from mono-hydrogen phosphate, carbonate, amide-III and CH_2 wag, respectively. We also observe the amide-I feature in the 1630-1656 cm^{-1} region. We calculated the mineral-to-matrix content as the phosphate $\nu 1$/amide I ratio, and phosphate-to-carbonate level (i.e. a marker of carbonate substitution) by the phosphate $\nu 1$/carbonate ratio. The degree of bone remodeling was estimated as the carbonate $\nu 1$/amide I ratio (since the carbonate-to-matrix ratio has been associated with increased risk of fracture), and mineral crystallinity as the reciprocal of full-width at half maximum (FWHM) of phosphate $\nu 1$ peak[16,21,23,26].
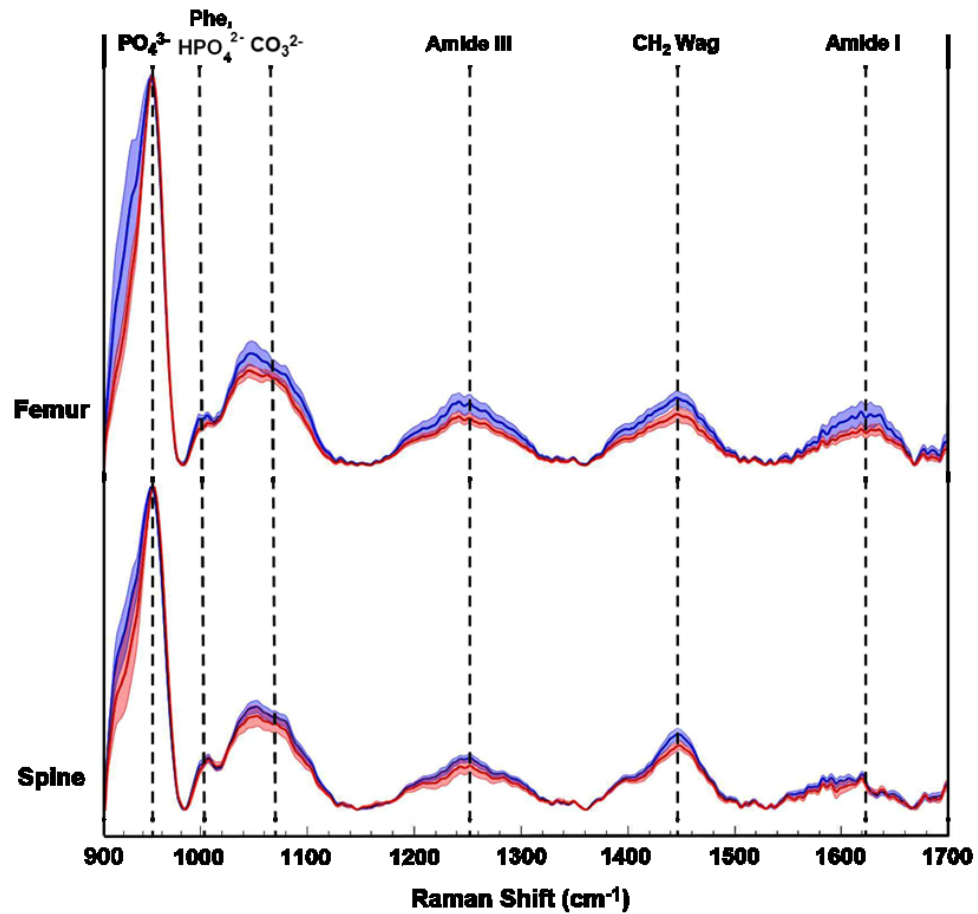
Figure 4.3 **Representative Raman spectra acquired from metastatic breast cancer affected femurs and spines.** Spectra (normalized to PO_4^{3-} v1 peak) were acquired from week 0 control group (blue tracings) and 5 weeks after tumor cell inoculations (red tracings). The solid lines depict the mean spectrum of each sample group with associated shadings representing the ±1 standard deviations (SD). Spectra are vertically offset for visualization purposes.

Figure 4.4A and **4.4B** show the (length-averaged) changes that occurred in these ratios in femurs and spines, respectively, during disease progression over the 5-week time course of the study. Relative to week 0, there was a progressive increase in the phosphate v1/amide I ratio (**Figure 4.4A & 4.4B**), as the tumor burden increased over time. This can be interpreted as being indicative of an increase in mineral density, which has previously been shown to be strongly negatively correlated to bone toughness[23]. Alternatively, within the context of this osteolytic metastatic model, we need to also consider the possibility that

both PO_4^{3-} and type I collagen are decreasing with the degradation of collagen, especially by week 4 & 5, surpassing the loss of PO_4^{3-} mineral content. Such a scenario reflects the fact that the colonizing tumor cells stimulate maturation of osteoclasts and hence bone resorption. This occurs by osteoclast-modulated acidification (~pH 4.5) of the resorption area that dissolves mineral and simultaneously releases of cathepsin K thereby degrading the organic bone matrix[40]. Stromal acidification would have also been enhanced by the proximal metabolic activities of the proliferating cancer cells, which generate low extracellular pH[40]. It has been shown that metabolic acidification causes large releases of bio-hydroxyapatite mineral from bone[41], and such acidification may contribute to and increase the loss in PO_4^{3-}. At the same time, stromal cell and tumor cell production of matrix metalloproteinases could be expected to extend matrix degradation beyond resorption sites[7,42,43]. This degradation of collagen has been recorded as blood borne type I collagen fragments, which are biomarkers of bone metastases[44]. In addition, over the time course of the study, one expects an increasing loss of collagen regeneration by osteoblasts because expression of their matrix generating genes are known to be inhibited by acidification. Moreover, osteoblasts undergo apoptosis during cancer colonization[28,29,41], which would curtail bone formation. Similar to the consideration that an increase in mineral density weakens bone, loss of both mineral and matrix components during increasing losses of collagen would be expected to compromise the mechanical properties of the bone[45].

Relative to week 0, changes in the carbonate substitution (PO_4^{3-} ν1/CO_3^{2-} ratio) do not exhibit a consistent trend. The ratio proceeds through an increase in CO_3^{2-} substitution at week 2, which is more pronounced in the femurs relative to the spine (compare **Figure 4.4A** to **4.4B**), and then to decreases at weeks 4 & 5 (**Figure 4.4A** & **4.4B**). The increase

at week 2 is consistent with previously reported results at a week 3 time point[16]. This type of substitution has been shown to decrease crystallinity size, which generates defects in the apatite matrix and weakens the bone[23]. The relatively high decrease in CO_3^{2-} substitutions by week 5 of metastatic progression likely reflects high levels of osteolytic activity, i.e., mineral dissolution, and concomitant utilization of released HCO_3^- by the rapidly growing tumor mass as a means to neutralize intracellular pH^{46}.

Throughout the course of metastatic progression in femurs as well as in spine, increased CO_3^{2-}/amide I ratios relative to week 0 were recorded (**Figure 4.4A & 4.4B**). Similar to our discussion for the phosphate-to-matrix ratio, the progressive increase in the CO_3^{2-}/amide I ratios in femurs possibly reflects unchecked degradation/loss of collagen that exceeded the rate of loss of CO_3^{2-}. Such increases in CO_3^{2-}/amide I ratios has been shown to be associated with increased fracture risk in the literature[47].
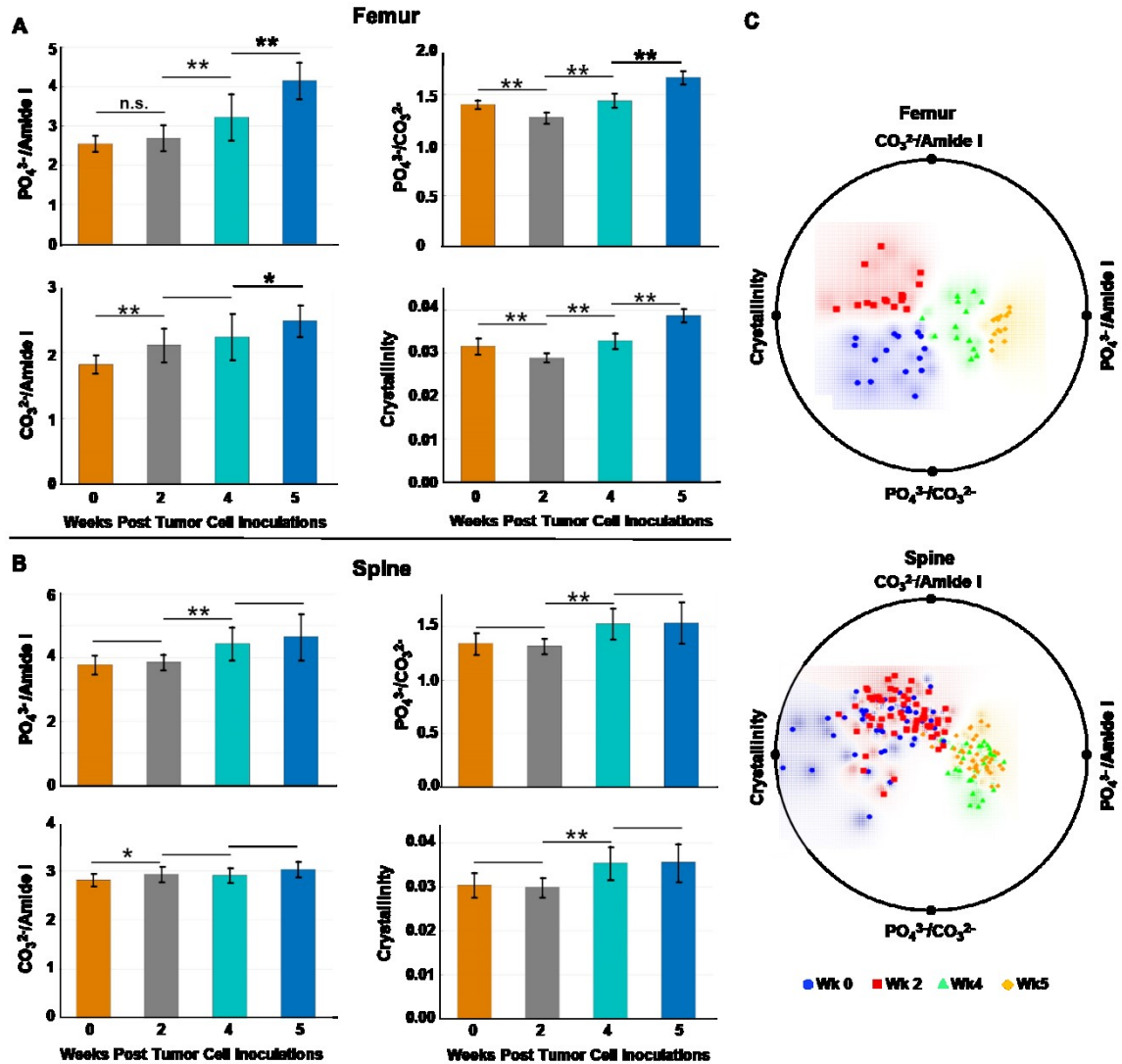
Figure 4.4 **Raman spectra-derived metrics of bone compositional changes at each week of the study and corresponding radial visualization plots.** Characteristics analyzed were: collagen mineralization as the PO_4^{3-}/amide I (phosphate v1/amide I) ratio, phosphate-to-carbonate ratio: PO_4^{3-}/CO_3^{2-} (phosphate v1/carbonate) ratio, remodelling as the CO_3^{2-}/amide I ratio, and mineral crystallinity from 1/FWHM PO_4^{3-} (1/FWHM phosphate v1 peak) calculations. Relative to week 0, average compositional changes of (A) femurs and (B) spines at 2, 4, and 5 weeks post tumor cell inoculations were quantified. Error bars = ±1 SD. (n,s, denotes not significant, *p<0.05, **p<0.005) (C) Distinct clustering of the spectral data corresponding to each week was revealed in the case of the femur analyses while two clusters emerged in the analysis of the spine data, namely, an early stage cluster: week 0 + week 2 and a late stage cluster: week 4 + week 5. Blue circles = week 0, red squares = week 2, green triangles = week 4, and orange diamond = week 5.

100

Relative to week 0, the bio-hydroxyapatite crystallinity (1/FWHM of PO_4^{3-} $\nu 1$)

decreased in femurs and was unchanged in spine at week 2 but thereafter increased in all

bone samples (**Figure 4.4A & 4.4B**). The decrease at week 2 is consistent with reported

results at a week 3 time point (16) and, as is generally the case, mirrors the increases in

CO_3^{2-} substitution (i.e., decreases in PO_4^{3-} $\nu 1$/CO_3^{2-} ratio) at this time point[48]. Replacement

of phosphate with carbonate at week 2 increases the number of defects in the bio-

hydroxyapatite lattice and reduces the crystallinity[25]. Likewise, increases of crystallinity at

the later time points parallels the decreases in carbonate content and may reflect an

anticipated accelerated bone turnover during osteolytic metastatic progression[48]. In

addition, increases in crystallinity is likely an indication of retarded bone formation with a

generation of larger thinner out-of-alignment crystallites rather than production of smaller

younger crystals resulting in increased fragility[17,48]. Increases of crystallinity have also

been previously reported at a week 4 time point for MDA-MB-435 breast cancer cell

metastases. While earlier time points were not evaluated, the authors suggested, given that

a metastatic bone prostate model revealed decreases in crystallinity[49], that this spectral

marker may differ with different primary tumors[26]. However, our data indicate that, at early

time points, decreased crystallinity can be associated with bone metastasis from breast

cancer, suggesting that both the progression of metastasis and primary tumor type may

impact changes in bone crystallinity.

We also performed radial visualization of the metrics generated from the femur and

spine to assess the degree of inter- vs intra-time point variability across the tested mice

specimen. **Figure 4.4C** shows distinct clustering patterns for the data points corresponding

to the femurs from each week of evaluation. Evidently, crystallinity and the PO_4^{3-}/amide I

ratios account for majority of the differential clustering of the femur data. For the spine measurements, we observed overlap between the week 0 and week 2 (early stage) as well as week 4 and week 5 (late stage) data points. Nevertheless, a clear distinction between the early and late stages of disease progression was revealed for both the femur and spine samples.

However, as metastasis to long bones is known to vary along their lengths[50], we reasoned that alterations in specific regions of the femurs may have been masked within the averaged values. **Figure 4.5A** indicates that changes in the computed metrics were more pronounced in the metaphysis of femurs and minimal changes in the spectral signatures/biochemical composition of diaphysis were registered. This is consistent with the known propensity of metastasizing tumor cells to colonize the highly vascularized remodeling niche of the metaphysis[9,26,28]. In contrast to the femur, **Figure 4.5B** indicates changes in bone compositional markers occur more uniformly throughout the length of vertebrae assessed. Yet, changes in bone composition were higher in vertebrae of the L5-S2 lumbar-sacral region, which is consistent with previous assessment of high tumor burden in these locations in this model system[27]. Therefore, including measurements of vertebrae below the pelvic region may skew the average values. Notably, the portion of spine analyzed in this model system broadly parallels clinical disease as the lumbar-sacral region is the second most frequent site of spinal compression due to metastatic disease in patients[51].
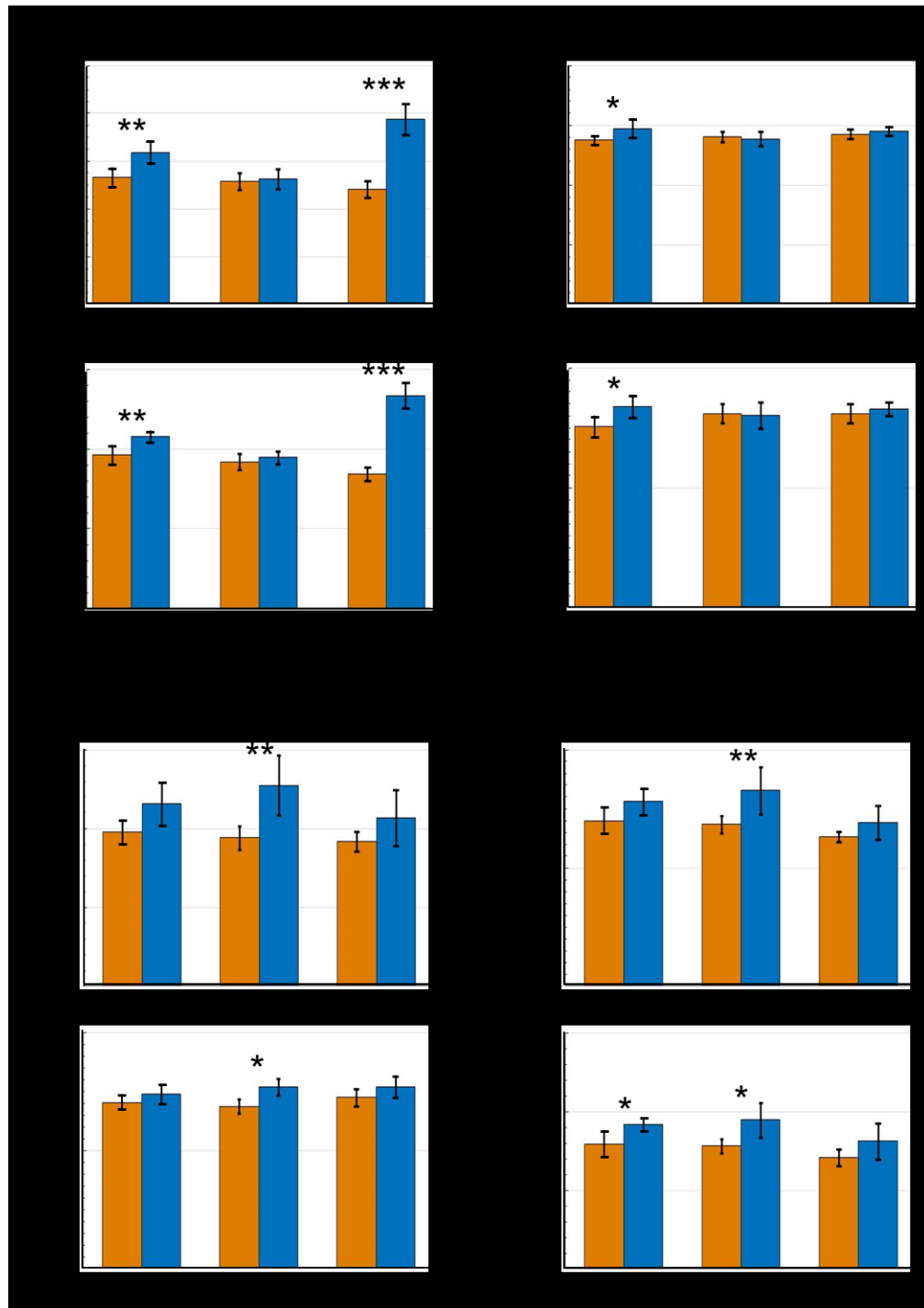
Figure 4.5 **Raman spectral-derived metrics of bone compositional changes as a function of the position of the measurements on the bone.** (A) Relative to week 0, average compositional changes (see **Figure 4.4**) at the distal metaphysis, diaphysis, and proximal metaphysis of femurs. (B) Relative to week 0, compositional changes at lumbar vertebrae (L1 – L4), lumbar – sacral vertebrae (L5 – S2), and sacral – caudal vertebrae (S3 – C2) of spines. Orange bar = week 0 and blue bar = week 4. Error bars = ± 1 SD. (* p<0.05, ** p<0.01, *** p<0.001).

To further assess the utility of the spectral information in recognizing progressive stages of metastatic bone alterations, SVM-derived decision algorithm was developed on the principal component (PC) scores. **Table 4.1A** shows classification accuracies of metastatic involvement at different regions of the femurs across the five weeks of study. The overall classification accuracy obtained for the SVM-derived decision algorithm was found to be 93.6% with the classification accuracy for any combination of site and time point being ≥83%. Expectedly, the classification of diaphysis exhibited the lowest accuracy when compared to that of the metaphyseal regions.

To ensure the robustness of this decision algorithm, we also conducted a negative control study where the spectra were randomly assigned to sites and time points, irrespective of their true identity[52]. In this control study, a maximum average accuracy of ca. 26% was achieved over 20 iterations underscoring the reliability of the actual SVM-derived decision model in discerning subtle, but consistent, metastasis driven changes during the 5-week course the study.

Table 4.1A Classification results for the SVM-derived decision algorithm as a function of time point and location in the femur.

Site	Week 0	Week 2	Week 4	Week 5	Average
Distal Metaphysis	100%	95%	95%	85%	93.8%
Diaphysis	100%	87%	83%	93%	90.8%
Proximal Metaphysis	100%	100%	95%	90%	96.3%

Table 4.1B Classification results for the SVM-derived decision algorithm using only the selected spectral features as a function of time point and location in the femur.

Site	Week 0	Week 2	Week 4	Week 5	Average
Distal Metaphysis	100%	100%	90%	90%	93.8%
Diaphysis	100%	83%	67%	76%	79.8%
Proximal Metaphysis	100%	100%	90%	70%	90.0%

Additionally, by restricting the analyses to only the prominent biomarkers noted above, we developed another decision algorithm to classify the femur sites at different time points (**Table 4.1B**). High classification accuracies were obtained for the analyses using selected Raman features, despite utilization of only 27% of the entire spectral information. In particular, the distal and proximal metaphysis largely maintained the same accuracy levels as the full spectrum decision model (**Table 4.1A**), while the misclassification rates were significantly higher for diaphysis. Also, as a maximum rate of osteolysis was approached (week 4-5), a slightly higher misclassification between week 4 and week 5 groups was obtained, especially for the proximal metaphysis. Corresponding classification results of spine are shown in supplementary information (**Table 4.2**). One notable feature in the classification analyses for the spine is the inability of the SVM decision model to accurately distinguish between week 0 and week 2 cases (that shows up as a reduction in classification accuracy at week 0, particularly for the caudal vertebras) indicating that metastasis-induced compositional changes may not occur as early as in the femurs.

Table 4.2A Classification results for the SVM-derived decision algorithm as a function of time point and location in the spine.

Site	Week 0	Week 2	Week 4	Week 5	Average
Lumbar vertebras	95%	100%	100%	100%	98.8%
Sacral vertebras	90%	100%	100%	90%	95.0%
Caudal vertebras	65%	100%	100%	100%	91.3%

Table 4.2B Classification results for the SVM-derived decision algorithm using only the selected spectral features as a function of time point and location in the spine.

Site	Week 0	Week 2	Week 4	Week 5	Average
Lumbar vertebras	85%	100%	95%	100%	95.0%
Sacral vertebras	90%	100%	100%	100%	97.5%
Caudal vertebras	65%	100%	100%	100%	91.3%

Once the Raman spectroscopic analyses were completed on femurs, validation of the findings was sought from fluorescence imaging within different regions of the femurs. As seen in **Figure 4.6A**, bright fluorescence from the tdTomato expressing tumor cells was readily recorded from metaphyseal regions of the femurs, especially at week 4 and 5, while diaphyseal regions showed no fluorescence signals. This is consistent with the results in **Figure 4.5A** as well as the previously reported preferential location of metastasis to the metaphyseal regions of femurs[28]. **Figure 4.6B** shows the assessment of the images in **Figure 4.6A** in terms of fold increases in fluorescence intensities across metaphysis as a function of metastatic progression. We observe that as early as week 2 fluorescence intensities from within the metaphyseal regions was about 2.5-fold higher than the autofluorescent signals captured at week 0, i.e., in femurs without tumor involvement. By week 4 the fluorescence intensity from the distal metaphysis was about *ca.* 3-4-fold higher than autofluorescence (week 0) while at the proximal metaphysis it was about 13-fold higher. A reverse trend was seen in the week 5 case where the fluorescence intensity was about 13-fold higher than background in the distal metaphysis while the proximal metaphysis harbored low amounts of tumor cells as reflected in a fluorescence intensity that was only about 1.5-2-fold higher than the autofluorescence of week 0. Except for comparisons between week 2 and week 4 distal metaphysis and week 2 and week 5 proximal metaphysis, fold increases in fluorescence intensities of each week of metastatic progression were all significant relative to week 0 as well as between weeks (two-tailed Student t-test, asterisk: $p < 0.05$). Overall, fluorescence imaging observations were consistent with the quantitative Raman spectroscopic analyses, and provided an understanding of the differential degree of bone deterioration as a function of increasing

metastatic tumor involvement over time. This imaging also reflects the heterogeneity and stochastic nature of metastatic progression, where, in this model system, metastasis to femur regions in different subjects can exhibit regional tumor burden differences; e.g., week 4 or 5 femurs with dissimilar metastatic burdens at different ends of the femurs.
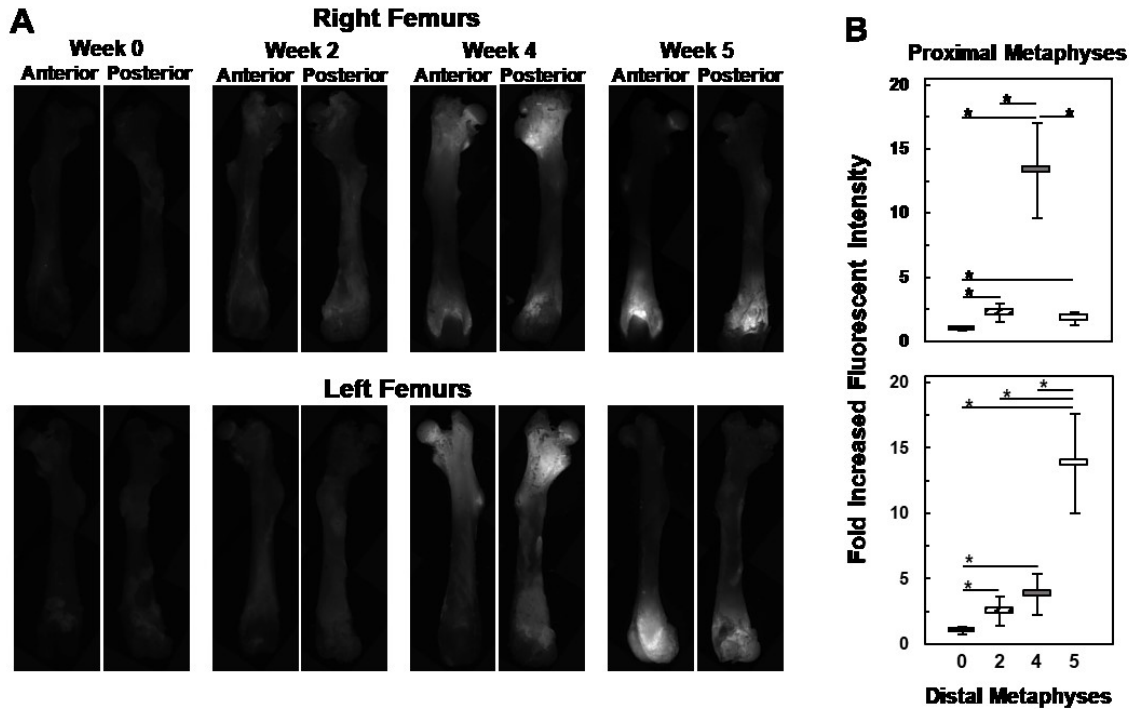
Figure 4.6 **Fluorescent imaging-based assessment of the metastatic lesions in femurs.** (A) Fluorescence images of anterior and posterior views of right (top panels) and left (bottom panels) femurs from each week (0, 2, 4, 5) of the study. Autofluorescence was low (week 0 images), metastasis specific fluorescent signals from tdT-435 cells within the metaphysis regions was relatively weak at week 2 and much more intense at week 4 and 5. (B) Fold increases in fluorescent intensities from the metaphysis regions of femurs in (A) relative to week 0 autofluorescence as well as between weeks as determined by the semi-quantitative measurements. Error bars = ±1 SD. Two-tailed Students t-test was employed for evaluating statistical significance (asterisk depicts p<0.05).

4.5 Conclusion

In this pilot study, we demonstrated that Raman spectroscopy has the capability to detect biochemical changes to the structure of bones associated with cancer metastasis without *a*

priori imaging or pathological knowledge of lesion location. Raman spectroscopy was able to identify sites of early tumor involvement, which points towards a potential diagnostic for early intervention not only to treat the advancing tumor but also to further assess these sites to circumvent pathological fractures. Hence, it may be possible to build an accurate risk assessment tool using Raman spectroscopy in conjunction with microCT or MRI. Along these lines, we showed that spectral changes emanating from the variations of specific spectral features can be utilized to construct decision algorithms with high diagnostic power.

Ultimately, we envision longitudinal studies will contribute to our understanding of molecular changes that indicate metastatic involvement at early stages. Movement towards bringing real time, fully noninvasive Raman spectroscopic assessments of bone to the clinic has been in progress in other laboratories, and impressive advances in instrumentation[53] and data processing[15] have been demonstrated. Together, such developments set the stage for future *in vivo* application of Raman spectroscopy for assessment of metastatic progression in bone and of fracture risk.

Acknowledgements

Bibliography

1. Siegel, R.L., Miller, K.D. and Jemal, A. Cancer Statistics, 2017. *CA Cancer J Clin* **2017**. DOI: 10.3322/caac.21387

2. DeSantis, C.E., Lin, C.C., Mariotto, A.B., Siegel, R.L., Stein, K.D., Kramer, J.L., et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J. Clin.* **2014**, 64(4):252-71. DOI: 10.3322/caac.21235

3. Berman, A.T., Thukral, A.D., Hwang, W.T., Solin, L.J. and Vapiwala, N. Incidence and patterns of distant metastases for patients with early-stage breast cancer after breast conservation treatment. *Clin. Breast Cancer* **2013**, 13(2):88-94. DOI: 10.1016/j.clbc.2012.11.001

4. Steinauer, K., Huang, D.J., Eppenberger-Castori, S., Amann, E. and Guth, U. Bone metastases in breast cancer: Frequency, metastatic pattern and non-systemic locoregional therapy. *J. Bone Oncol.* **2014**, 3(2):54-60. DOI: 10.1016/j.jbo.2014.05.001

5. Oster, G., Lamerato, L., Glass, A.G., Richert-Boe, K.E., Lopez, A., Chung, K., et al. Natural history of skeletal-related events in patients with breast, lung, or prostate cancer and metastases to bone: a 15-year study in two large US health systems. *Support Care Cancer* **2013**, 21(12):3279-86. DOI: 10.1007/s00520-013-1887-3

6. Fontanella, C., Fanotto, V., Rihawi, K., Aprile, G. and Puglisi, F. Skeletal metastases from breast cancer: pathogenesis of bone tropism and treatment strategy. *Clin. Exp. Metastasis* **2015**, 32(8):819-33. DOI: 10.1007/s10585-015-9743-0

7. Liu, B., Cui, J., Sun, J., Li, J., Han, X., Guo, J., et al. Immunolocalization of MMP9 and MMP2 in osteolytic metastasis originating from MDA-MB-231 human breast cancer cells. *Mol. Med. Rep.* **2016**, 14(2):1099-106. DOI: 10.3892/mmr.2016.5374

8. Lin, J., Goldstein, L., Nesbit, A. and Chen, M.Y. Influence of Hormone Receptor Status on Spinal Metastatic Lesions in Patients with Breast Cancer. *World Neurosurg* **2016**, 85:42-8. DOI: 10.1016/j.wneu.2015.07.068

9. Märdian, S., Schaser, K.D., Ruppert, M., Melcher, I., Haas, N.P. and Schwabe, P. Skeletal metastatic disease of the femur: results by management with intramedullary nailing. *Acta. Chir. Orthop. Traumatol. Cech.* **2015**, 82(3):192-7.

10. Jambor, I., Kuisma, A., Ramadan, S., Huovinen, R., Sandell, M., Kajander, S., et al. Prospective evaluation of planar bone scintigraphy, SPECT, SPECT/CT, 18F-NaF PET/CT and whole body 1.5T MRI, including DWI, for the detection of bone metastases in high risk breast and prostate cancer patients: SKELETA clinical trial. *Acta. Oncol.* **2016**, 55(1):59-67. DOI: 10.3109/0284186X.2015.1027411

11. Kannivelu, A., Loke, K.S., Kok, T.Y., Osmany, S.Y., Ali, S.Z., Suat-Jin, L., et al. The role of PET/CT in the evaluation of skeletal metastases. *Semin. Musculoskelet Radiol.* **2014**, 18(2):149-65. DOI: 10.1055/s-0034-1371017

12. Damron, T.A., Nazarian, A., Entezari, V., Brown, C., Grant, W., Calderon, N., et al. CT-based Structural Rigidity Analysis Is More Accurate Than Mirels Scoring for Fracture Prediction in Metastatic Femoral Lesions. *Clin. Orthop. Relat. Res.* **2016**, 474(3):643-51. DOI: 10.1007/s11999-015-4453-0

13. Bolotin, H.H. DXA *in vivo* BMD methodology: an erroneous and misleading research and clinical gauge of bone mineral status, bone fragility, and bone remodelling. *Bone* **2007**, 41(1):138-54. DOI: 10.1016/j.bone.2007.02.022

14. Snyder, B.D., Cordio, M.A., Nazarian, A., Kwak, S.D., Chang, D.J., Entezari, V., et al. Noninvasive Prediction of Fracture Risk in Patients with Metastatic Cancer to the Spine. *Clin. Cancer Res.* **2009**, 15(24):7676-83. DOI: 10.1158/1078-0432.CCR-09-0420

15. Buckley, K., Kerns, J.G., Vinton, J., Gikas, P.D., Smith, C., Parker, A.W., et al. Towards the *in vivo* prediction of fragility fractures with Raman spectroscopy. *J. Raman Spectrosc.* **2015**, 46(7):610-18. DOI: 10.1002/jrs.4706

16. Burke, M.V., Atkins, A., Akens, M., Willett, T.L. and Whyne, C.M. Osteolytic and mixed cancer metastasis modulates collagen and mineral parameters within rat vertebral bone matrix. *J. Orthop. Res.* **2016**, 34(12):2126-36. DOI: 10.1002/jor.23248

17. Paschalis, E.P., Gamsjaeger, S., Hassler, N., Klaushofer, K. and Burr, D. Ovarian hormone depletion affects cortical bone quality differently on different skeletal envelopes. *Bone* **2017**, 95:55-64. DOI: 10.1016/j.bone.2016.10.029

18. Unal, M. and Akkus, O. Raman spectral classification of mineral- and collagen-bound water's associations to elastic and post-yield mechanical properties of cortical bone. *Bone* **2015**, 81:315-26. DOI: 10.1016/j.bone.2015.07.024

19. Feng, X., Moy, A.J., Markey, M.K., Fox, M.C., Reichenberg, J.S. and Tunnell, J.W. Biophysical basis for noninvasive skin cancer detection using Raman spectroscopy. *SPIE BiOS* **2016**, pp. 97040C-97040C. DOI: 10.1117/12.2209421

20. Mahadevan-Jansen, A. and Richards-Kortum, R.R. Raman spectroscopy for the detection of cancers and precancers. *Journal of Biomedical Optics* **1996**, 1(1):31-70. DOI: 10.1117/12.227815

21. Mandair, G.S. and Morris, M.D. Contributions of Raman spectroscopy to the understanding of bone strength. *Bonekey Rep.* **2015**, 4:620. DOI: 10.1038/bonekey.2014.115

22. Bi, X., Patil, C.A., Lynch, C.C., Pharr, G.M., Mahadevan-Jansen, A. and Nyman, J.S. Raman and mechanical properties correlate at whole bone- and tissue-levels in a genetic mouse model. *J. Biomech.* **2011**, 44(2):297-303. DOI: 10.1016/j.jbiomech.2010.10.009

23. Inzana, J.A., Maher, J.R., Takahata, M., Schwarz, E.M., Berger, A.J. and Awad, H.A. Bone fragility beyond strength and mineral density: Raman spectroscopy predicts femoral fracture toughness in a murine model of rheumatoid arthritis. *J. Biomech.* **2013**, 46(4):723-30. DOI: 10.1016/j.jbiomech.2012.11.039

24. McNerny, E.M., Gong, B., Morris, M.D. and Kohn, D.H. Bone fracture toughness and strength correlate with collagen cross-link maturity in a dose-controlled lathyrism mouse model. *J. Bone Miner Res.* **2015**, 30(3):455-64. DOI: 10.1002/jbmr.2356

25. Bi, X., Sterling, J.A., Merkel, A.R., Perrien, D.S., Nyman, J.S. and Mahadevan-Jansen, A. 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* **2013**, 56(2):454-60. DOI: 10.1016/j.bone.2013.07.006

26. Ding, H., Nyman, J.S., Sterling, J.A., Perrien, D.S., Mahadevan-Jansen, A. and Bi, X. Development of Raman spectral markers to assess metastatic bone in breast cancer. J *Biomed. Opt.* **2014**, 19(11):111606. DOI: 10.1117/1.JBO.19.11.111606

27. Harms, J.F. and Welch, D.R. MDA-MB-435 human breast carcinoma metastasis to bone. *Clin. Exp. Metastasis* **2003**, 20(4):327-34. DOI: 10.1023/A:1024062911144

28. Phadke, P.A., Mercer, R.R., Harms, J.F., Jia, Y., Frost, A.R., Jewell, J.L., et al. Kinetics of metastatic breast cancer cell trafficking in bone. *Clin. Cancer Res.* **2006**, 12(5):1431-40. DOI: 10.1158/1078-0432.CCR-05-1806

29. Mastro, A.M., Gay, C.V., Welch, D.R., Donahue, H.J., Jewell, J., Mercer, R., et al. Breast cancer cells induce osteoblast apoptosis: a possible contributor to bone degradation. *J. Cell Biochem.* **2004**, 91(2):265-76. DOI: 10.1002/jcb.10746

30. Matousek, P. and Stone, N. Development of deep subsurface Raman spectroscopy for medical diagnosis and disease monitoring. *Chem. Soc. Rev.* **2016**, 45(7):1794-802. DOI: 10.1039/c5cs00466g

31. Brinkley, B.R., Beall, P.T., Wible, L.J., Mace, M.L., Turner, D.S. and Cailleau, R.M. Variations in cell form and cytoskeleton in human breast carcinoma cells in vitro. *Cancer Res.* **1980**, 40(9):3118-29.

32. Winnard, P.T., Jr., Kluth, J.B. and Raman, V. Noninvasive optical tracking of red fluorescent protein-expressing cancer cells in a model of metastatic breast cancer. *Neoplasia* **2006**, 8(10):796-806. DOI:10.1593/neo.06304

33. Paidi, S.K., Siddhanta, S., Strouse, R., McGivney, J.B., Larkin, C. and Barman, I. Rapid Identification of Biotherapeutics with Label-Free Raman Spectroscopy. *Anal. Chem.* **2016**, 88(8):4361-8. DOI: 10.1021/acs.analchem.5b04794

34. Beier, B.D. and Berger, A.J. Method for automated background subtraction from Raman spectra containing known contaminants. *Analyst* **2009**, 134(6):1198-202. DOI: 10.1039/b821856k

35. Ringner, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, 26(3):303-4. DOI: 10.1038/nbt0308-303

36. Genton, M.G. Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research* **2001**, 2:299-312

37. Suykens, J.A., Van Gestel, T., Vandewalle, J. and De Moor, B. A support vector machine formulation to PCA analysis and its kernel version. *IEEE Transactions on neural networks* **2003**, 14(2):447-50. DOI: 10.1109/TNN.2003.809414

38. Pelckmans, K., Suykens, J.A., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., et al. LS-SVMlab: a matlab/c toolbox for least squares support vector machines. Tutorial. *KULeuven-ESAT.* Leuven, Belgium. **2002**, 142:1-2

39. Rubin, M.R., Paschalis, E.P., Poundarik, A., Sroga, G.E., McMahon, D.J., Gamsjaeger, S., et al. Advanced Glycation Endproducts and Bone Material Properties in Type 1 Diabetic Mice. *PLoS One* **2016**, 11(5):e0154700. DOI: 10.1371/journal.pone.0154700

40. Yoneda, T., Hiasa, M., Nagata, Y., Okui, T. and White, F.A. Acidic microenvironment and bone pain in cancer-colonized bone. *Bonekey Rep.* **2015**, 4:690. DOI: 10.1038/bonekey.2015.58

41. Bushinsky, D.A., Chabala, J.M., Gavrilov, K.L. and Levi-Setti, R. Effects of *in vivo* metabolic acidosis on midcortical bone ion composition. *Am. J. Physiol.* **1999**, 277(5):F813-9.

42. Dutta, A., Li, J., Lu, H., Akech, J., Pratap, J., Wang, T., et al. Integrin alphavbeta6 promotes an osteolytic program in cancer cells by upregulating MMP2. *Cancer Res.* **2014**, 74(5):1598-608. DOI: 10.1158/0008-5472.CAN-13-1796

43. Shimamura, T., Amizuka, N., Li, M., Freitas, P.H., White, J.H., Henderson, J.E., et al. Histological observations on the microenvironment of osteolytic bone metastasis by

breast carcinoma cell line. *Biomed. Res.* **2005**, 26(4):159-72. DOI: 10.2220/biomedres.26.159

44. Dean-Colomb, W., Hess, K.R., Young, E., Gornet, T.G., Handy, B.C., Moulder, S.L., et al. Elevated serum P1NP predicts development of bone metastasis and survival in early-stage breast cancer. *Breast Cancer Res. Treat* **2013**, 137(2):631-6. DOI: 10.1007/s10549-012-2374-0

45. Depalle, B., Qin, Z., Shefelbine, S.J. and Buehler, M.J. Influence of cross-link structure, density and mechanical properties in the mesoscale deformation mechanisms of collagen fibrils. *J. Mech. Behav. Biomed. Mater.* **2015**, 52:1-13. DOI: 10.1016/j.jmbbm.2014.07.008

46. Lee, S., Mele, M., Vahl, P., Christiansen, P.M., Jensen, V.E. and Boedtkjer, E. Na^+, HCO_3^- -cotransport is functionally upregulated during human breast carcinogenesis and required for the inverted pH gradient across the plasma membrane. *Pflugers Arch.* **2015**, 467(2):367-77. DOI: 10.1007/s00424-014-1524-0

47. Boskey, A.L. Bone composition: relationship to bone fragility and antiosteoporotic drug effects. *Bonekey Rep.* **2013**, 2:447. DOI: 10.1038/bonekey.2013.181

48. Gong, B., Oest, M.E., Mann, K.A., Damron, T.A. and Morris, M.D. Raman spectroscopy demonstrates prolonged alteration of bone chemical composition following extremity localized irradiation. *Bone* **2013**, 57(1):252-8. DOI: 10.1016/j.bone.2013.08.014

49. Esmonde-White, K.A., Sottnik, J., Morris, M. and Keller, E. Raman spectroscopy of bone metastasis. *Photonic Therapeutics and Diagnostics VIII* **2012**, 8207:82076P. DOI: 10.1117/12.909327

50. Lynch, M.E. and Fischbach, C. Biomechanical forces in the skeleton and their relevance to bone metastasis: biology and engineering considerations. *Advanced drug delivery reviews* **2014**, 79:119-34. DOI: 10.1016/j.addr.2014.08.009

51. Prasad, D. and Schiff, D. Malignant spinal-cord compression. *Lancet Oncol.* **2005**, 6(1):15-24. DOI: 10.1016/S1470-2045(04)01709-7

52. Winnard, P.T., Zhang, C., Vesuna, F., Kang, J.W., Garry, J., Dasari, R.R., et al. Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes. *Oncotarget* **2017**, 8(12):20266-87. DOI: 10.18632/oncotarget.14865

53. Demers, J.L., Esmonde-White, F.W., Esmonde-White, K.A., Morris, M.D. and Pogue, B.W. Next-generation Raman tomography instrument for non-invasive *in vivo* bone imaging. *Biomed. Opt. Express* **2015**, 6(3):793-806. DOI: 10.1364/BOE.6.000793

Chapter 5. Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes

5.1 Abstract

Molecular characterization of organ-specific metastatic lesions, which distinguish them from the primary tumor, will provide a better understanding of tissue specific adaptations that regulate metastatic progression. Using an orthotopic xenograft model, we have isolated isogenic metastatic human breast cancer (IMBC) cell lines directly from organ explants that are phenotypically distinct from the primary tumor cell line. Label-free Raman spectroscopy was used, and informative spectral bands were ascertained as differentiators of organ-specific metastases as opposed to the presence of a single universal marker. Decision algorithms derived from the Raman spectra unambiguously identified these isogenic cell lines as unique biological entities – a finding reinforced through metabolomic analyses that indicated tissue of origin metabolite distinctions between the cell lines. Notably, complementarity of the metabolomics and Raman datasets was found. Our findings provide evidence that metastatic spread generates tissue-specific adaptations at the molecular level within cancer cells, which can be differentiated with Raman spectroscopy.

5.2 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]. The American Cancer Society recently reporting a 5-year survival rate near 99% for local breast cancer[1,2]. However, the 5-year survival for metastatic breast cancer that involves distant organs drops to a dismal 24%[1,2]. This situation persists because understanding the metastatic progression of breast cancer remains challenging. This is due to several factors including a limited predictability as to which primary tumor is prone to metastatic progression, an inability to monitor the onset of successful metastatic growth, and incomplete knowledge of metabolic, physiologic, and molecular adaptations that allow for the cancer to survive and thrive within the different tissue types[3]. As such, procuring safe and efficacious chemotherapeutic regimen strategies that ablate metastatic lesions is an unmet clinical need[4]. In addition, the current practice of systemic administration of cytotoxic chemotherapy is limited with respect to targeting and drug resistance, which results in numerous adverse side-effects and no cures[5].

When considering potential solutions to this problem an important factor is the divergence of the metastatic cancer cells growing in visceral organs from the primary breast tumor cells[6-18]. Thus, there is a growing consensus from retrospective as well as prospective clinical trials that matched primary breast tumor and metastatic lesion biopsy samples often exhibit divergent expression of established biomarkers, for example, ER and HER2[7,9-11,17]. Therefore, metastatic lesions should not be considered simply as primary tumor implants at new sites but instead as significantly divergent tissue-specific lesions, which reflect adaptations to organ-specific environments[18]. Importantly, it is very difficult to discern if

various organ-specific metastatic lesions will have similar sensitivity to prescribed therapeutic regimens. Accordingly, the organ-specificity of the metastatic spread needs critical reconsideration, as, at present, databases of molecular profiles of matched primary and metastatic breast tumors have not been compiled to address global distinctions between metastatic sites and thus, cannot facilitate generalized metastatic site-specific nor patient specific smart therapeutic alternatives. Consequently, present clinical treatment decision options for distant metastatic breast cancer that rely on an evaluation of a few select biomarkers found during assessment of the primary tumor, although beneficial to subpopulations of patients[19], is also a likely contributing factor to the overall diminished response rates for survival from metastatic disease[9-11,13]. Such a conclusion is in line with the reported presence of altered and distinct biomarker signatures of the metastatic lesions with respect to those found at the corresponding primary tumor[7,9-11,17,18], which, when evaluated, may indicate that a change in an ongoing treatment strategy should be considered.

Dissecting metastatic cancers based on objective molecular markers remains an important challenge. Here, we propose a fundamentally different approach towards identification of defining metastatic cancer cell signatures from those of primary tumor cells. Harnessing the exquisite specificity of Raman spectroscopy in detecting molecular phenotypes in cells and tissue, we aimed to obtain rapid and label-free profiling of newly generated isogenic metastatic human breast cancer (IMBC) cell lines, which were produced from a xenograft mouse model. 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[20]. While this spectroscopic

technique has been recently used to distinguish between, normal, benign, and malignant breast tissues, by us and others[21-24], the potential for using these spectral markers as new routes to recognition of metastatic cell types that are isogenic to the primary tumor, as is the clinical case, has been understudied.

Starting from an orthotopic xenograft-based mouse model system, the studied human cell lines were obtained from cultured organ-of-origin explants of brain, liver, lung, and spine, as well as from the primary, i.e., mammary fat pad (MFP), site. These metastatic sites are representative of the common clinically observed breast cancer metastatic destinations[25,26] with spine representative of bone. Despite being isogenic, these cell lines exhibit important morphological and growth distinctions that support our hypothesis that each metastatic site imbues metastatic tumors with unique molecular attributes. Our Raman spectroscopic measurements reveal the presence of consistent spectral differences of the cell lines. Using multivariate chemometric methods, we show that these spectral changes can be utilized to develop decision algorithms with high diagnostic power. Furthermore, we identify the presence of spectrally informative features that bring to light each cell line's unique spectral characteristics, which reflect the inherent biochemical distinctions. We reason that these differences are a result of intricate reciprocal interactions between the cancer cells, parenchyma, and stroma at the target organ during metastatic growth. Combined with the ability to assay the stromal features, our findings underscore the relevance of Raman spectral information in characterizing isogenic metastatic lesions at different sites in terms of inherent biochemical determinants without staining or requiring *a priori* knowledge of the molecular transformations. In addition, preliminary metabolomic analyses provide supporting data indicating that cancer cells from different metastatic sites

acquire metabolic changes, which may define a cell line's metastatic organ of origin. Notably, a complementary overlap between the metabolite distinction data set and Raman spectroscopic signatures was also found.

5.3 Experimental section

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 and initial weights of about 19–20 g, 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. The human breast cancer cell line, MDA-MB-435, was obtained from ATCC. The MDA-MB-435 cell line was established in 1976 from a pleural effusion from an untreated 31-year-old female diagnosed with adenocarcinoma of the breast[27,28]. MDA-MB-435 cells were authenticated at the Johns Hopkins Genetic Resource Core Facility with the short tandem repeat marker results cross checked against cell lines at the ATCC bank. Generation and characterization of the parental MDA-MB-435-tdTomato (435-tdT) cell line has been previously described[29.] All culturing was done in standard humidified incubators at 37 °C and 5% CO_2. Primary tumors were initiated by injection of 2×10^6 435-tdT cells into the second thoracic mammary fat pad of 5 female NOD-SCID mice. After 13 - 15 weeks of primary tumor growth the mice were sacrificed and primary tumor, brain, liver, lungs, and, spine, were immediately excised from individual animals, dissected away from fat and muscle, and placed into sterile PBS on ice.

Pieces of primary tumor, and heavily diseased lungs, and a 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. In all cases, tissues from individual animals were cultured separately and there was no pooling of tissue samples.

All organ/bone tissue explants were initially cultured in RPMI-10% FBS supplemented with antibiotics (100 I.U./ml penicillin, 100 mg/ml streptomycin, 100 mg/ml ampicillin, and 100 mg/ml kanamycin) and, as necessary, Fungizone. The latter was often used during culturing cells out of spine as these pieces of bone, 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. Over the course of 1-3 months pure red fluorescent cell cultures were obtained and the use of pen/strep in the medium was eliminated. During routine passages the medium/floating cells was first collected and the adherent colonies were then lifted off the plates by room temperature incubations in HANKS-5 mM EDTA solution for ~5-10 min with shaking and tapping by hand. Lifted cells were pooled with the collected medium/cells, centrifuged 200 xg at 21 °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 12 hr to 24

hr and at times 48 hr (generally during recovery from -80 °C storage) for the larger percentage of adherent cells to settle and start to grow.

Average specific growth rate and average length of cell cycle division. Growth rate analyses were initiated by seeding 24 well plates with 105 cells per well and harvesting quadruplicates of these wells every 24 hr through to the 144 hr end-point. Growth curves were generated from live cell counts obtained with a TC10 Automatic Cell Counter (Bio-Rad) in the presence of Trypan Blue. Average specific growth rates[30]: μ, were obtained from the slopes of plots of $\ln(N_t/N_o)$ versus time, i.e., $\ln(N_t/N_o) = \mu t$, where N_t is the number of cells at time 't', N_o is the initial number of cells, and t is time. Consequently, the average length of the cell cycle was obtained from the equation: $t_c = \ln2/\mu$[30]. The rationale for the time intervals given in **Table 5.1** is: The initial number of cells: N_o, can only be obtained after the cells have had time to settle, adhere, and begin to grow, i.e., 1 day after seeding the cells.

Table 5.1 Average specific growth rates and length of cell cycle divisions for the indicated growth periods.

		μ' (Day-1)	μ (hr-1)	t_c[‡] (Days)	t_c (hr)
Brain	24-72 hr	0.149	0.0062	4.65	111.8
	72-144 hr	0.495	0.0206	1.40	33.6
Parental	24-144 hr	0.383	0.0160	1.81	43.4
Liver	24-120 hr	0.430	0.0179	1.61	38.2
	96-144 hr	0.076	0.0032	9.08	218.0
Lung	24-72 hr	0.555	0.0231	1.25	30.0
	72-144 hr	0.354	0.0148	1.96	47.0
Spine	24-120 hr	0.464	0.0193	1.49	35.8
	96-144 hr	0.062	0.0026	11.18	268.3
1º Tumor	24-72 hr	0.613	0.0255	1.13	27.2
	72-144 hr	0.142	0.0059	4.89	117.3

† Denotes average specific growth rate from equation: $\ln(N_t/N_o) = \mu t$, where N_t is the number of cells at time 't', N_o is the initial number of cells, and t is time.
‡ Denotes the average length of the cell cycle division and is derived from: $t_c = \ln2/\mu$.
1° Tumor denotes primary tumor.

Motility assay. Standard motility assays were done in 24 well Transwell® plates (Costar) with 8.0 μm membrane inserts. Cells were seeded into duplicate upper chambers at a density of 10,000 cells/well in 200 ml of RPMI-0.1% FBS medium while lower chambers contained 500 ml of RPMI-5% FBS medium. Cells at the bottom surface of membranes were counted daily using a 10× objective on an inverted fluorescence microscope (Nikon Eclipse TS100) with the inherent red fluorescence of tdT as a visual marker. Two separate experiments were done and for each experiment two fields of view were counted from each well. Results indicate means ± 1 standard deviation.

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.

Metabolomics: principle component analysis and heat map generation. Metabolite data from all samples were acquired using Agilent 6540 Quadrupole–Time-of-Flight (Q-TOF) mass spectrometer with Agilent 1290 HPLC at the Metabolomics Facility at Johns Hopkins Medical Institutions. Data was analyzed using Agilent Mass Hunter and Agilent Mass Profiler Professional (MPP) version 13.1.1 and Agilent Qualitative and Quantitative Analysis Software packages (version 6.00) to determine the metabolic profile of each sample.

Principal component analysis (PCA) was performed to study similarities and differences among the different samples. It is a linear transformation used to describe high dimensional data[31,32]. Expression values of metabolites and lipids were analyzed on Partek Genomics Studio 6.6 (Partek, Inc.) and used to create PCA plots. Each sphere represents a sample and each axis represents the principal components with the largest contributors being displayed. The distance between samples is inversely related to the similarity of their expression profiles, thus closely clustered samples are closely correlated. Hierarchical clustering was used to group similar expression patterns into clusters, which produced dendrograms that display the hierarchy of clustering. We clustered rows (expression values) and columns (samples) based on Euclidean distance and used average linkage method.

Raman spectroscopy. The custom-built Raman microscope (**Appendix Figure S5.1**) used in this work was previously reported[21]. 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. 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. Spectra (100: 10 × 10) were collected from 90 μm × 90 μm areas 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 spectroscopy 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 background corrected, normalized for intensity variations, and subsequently 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[33]. 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, we 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)[34].

The number of loading vectors (LV) used in the PLS-DA model was determined based on the minimal misclassification rate in a LOOCV 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.

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. We 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[22,35].

5.4 Results

Isogenic metastatic breast cancer cell lines from specific organs. In order to facilitate the tracking of metastatic progression in live mice[29], we engineered triple negative MDA-MB-435 human breast cancer cells[36-42] to stably express a red fluorescence protein (tdTomato) and here designates this cell line: 435-tdT. Using 435-tdT cells, we initiated the culturing of new organ specific metastatic breast cancer cells (**Figure 5.1**) with the inoculation of 435-tdT cells into the second thoracic mammary fat pads of female NOD-SCID mice. Phase-contrast images of fresh organ explants showed unresolved amorphous material without evidence of metastatic lesions while the bright tdT-fluorescence revealed the presence of the cancer (**Figure 5.1A**). Identified metastatic lesions were placed into cell culture and metastatic cancer cells grew out of native tissue environments until pure populations of red fluorescent cancer cells were obtained.
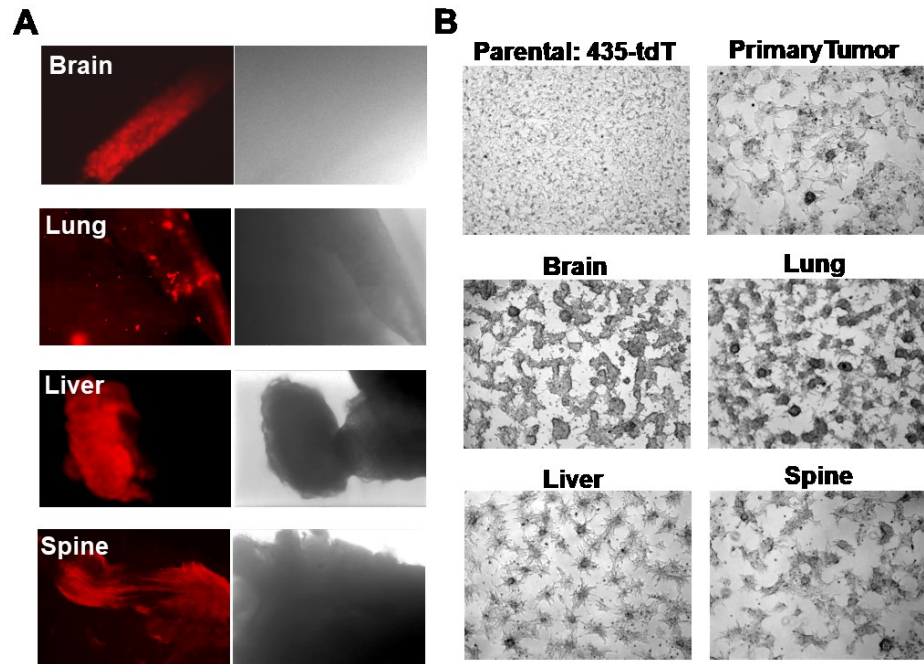
Figure 5.1 **Use of fluorescent microscopy to assess the locations of metastatic lesions in ex vivo organ samples and the growth patterns of the subsequent pure metastatic cell lines.** (A) Fluorescence and corresponding phase-contrast images of brain, lung, liver, and spine tissue explants immediately after dissection. (B) Phase contrast images of the different colony growth patterns of pure brain, liver, lung, and spine metastatic sublines as well as the primary tumor cell line, compared to the monolayer growth pattern of parental 435-tdT cells. Scale bars in all images depict 100 μm.

Once adapted to plastic, all metastatic cell lines as well as the primary tumor cells grew as loosely adherent 3D spherical colonies made up of tightly packed spherical cells with various degrees of monolayer growth (**Figure 5.1B**), which is starkly different from the overall monolayer growth of the parental 435-tdT cell line (**Figure 5.1B** and **Figure 5.2A** and **5.2B**).
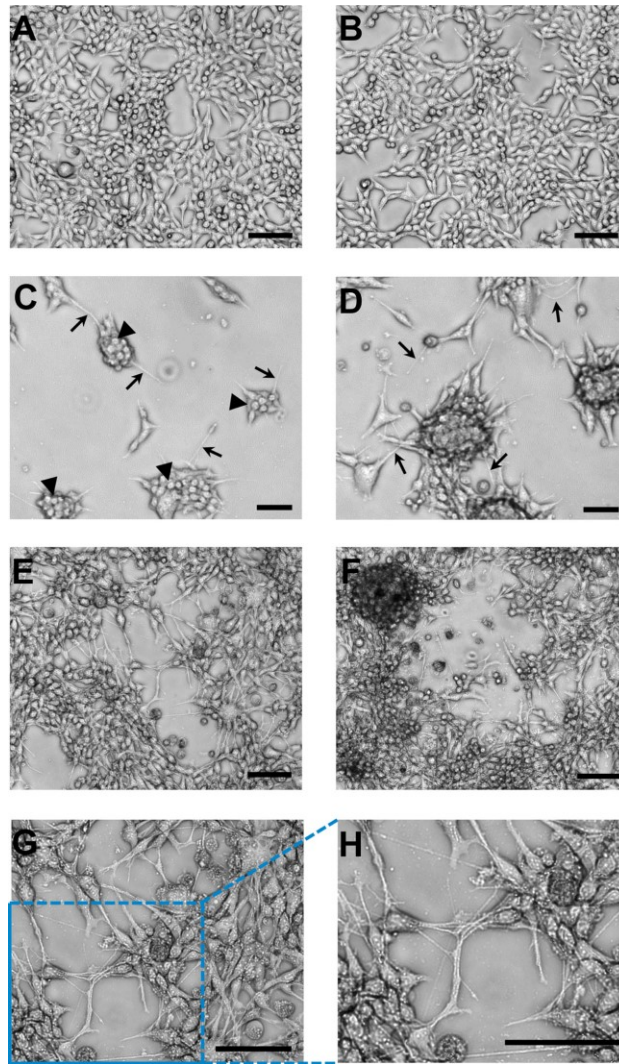
Figure 5.2 **Representative images of the brain cell line growth patterns on adherent plastic compared to monolayer growth of the parental cell line.** (A–B) Two fields-of-view of characteristic monolayer growth of the parental cell line. (C) Distinct separate colony growth was apparent at 48 hr post inoculation of the plate with distinct small spherical cells making up each colony (arrow heads) and thin cellular extensions/filopodia (micro- or nanotubes; arrows). (D) After 120 hr the interconnected colony pattern remained. (E–F) Two examples of the characteristic growth pattern at "confluency" of the brain cell line with colonies elaborately linked together by nanotubes. These interconnections between cells/colonies have consistently been recorded at > 100 μm in length. (G) Higher magnification of the central portion of image (E). (H) Expanded image of the lower left-hand corner of image (G). These magnified images allow for a very clear visualization of the complex and intricate web of interconnections between colonies that were in place. Scale bars in all images depict 100 μm.

130

Representative images exemplified by brain cell line growth patterns show the brain cell line had a colony growth pattern as contrasted to the monolayer growth of the parental cell line (**Figure 5.2** and **5.3**). Colonies were apparent as early as 24-48 hr post seeding of (**Figure 5.2C–D** and **Figure 5.3A**-top and middle images). Distinct small spherical cells making up each colony are readily seen (**Figure 5.2C**) and this growth pattern remains throughout culturing as exemplified at 120 hr of growth (**Figure 5.3A**-bottom images). In addition, thin cellular extensions, micro- or nanotubes, (**Figure 5.2C-D** and **Figure 5.3A**) were visible, which at relatively low cell numbers, i.e., at 24–48 hr, appeared to be attached to the substratum and also as connections between adjacent colonies. These connections become more numerous as the cultures grew (**Figure 5.2D** and **Figure 5.3A**). A characteristic pattern at "confluency" (**Figure 5.2E-F**) with colonies elaborately linked together by nanotubes is contrasted to monolayer growth patterns of the parental cell line (**Figure 5.2A-B**). Interconnections between cells/colonies have consistently been recorded at > 100 μm in length (**Figure 5.2C-H** and **Figure 5.3A**) when connecting distant cells/colonies. At high cell/colony numbers, depicted in two fields-of-view (**Figure 5.2E-F**) and in magnified and expanded images (**Figure 5.2G-H**), show that a complex and intricate web of interconnections between colonies often occurred. In addition, free floating small and large mammospheres (**Figure 5.3B**) and small floating colonies along with free floating single cells (most abundant in brain and spine cell cultures) were a consistent feature of these cultures, which was unexpected under the adherent plate conditions used throughout the culturing process.
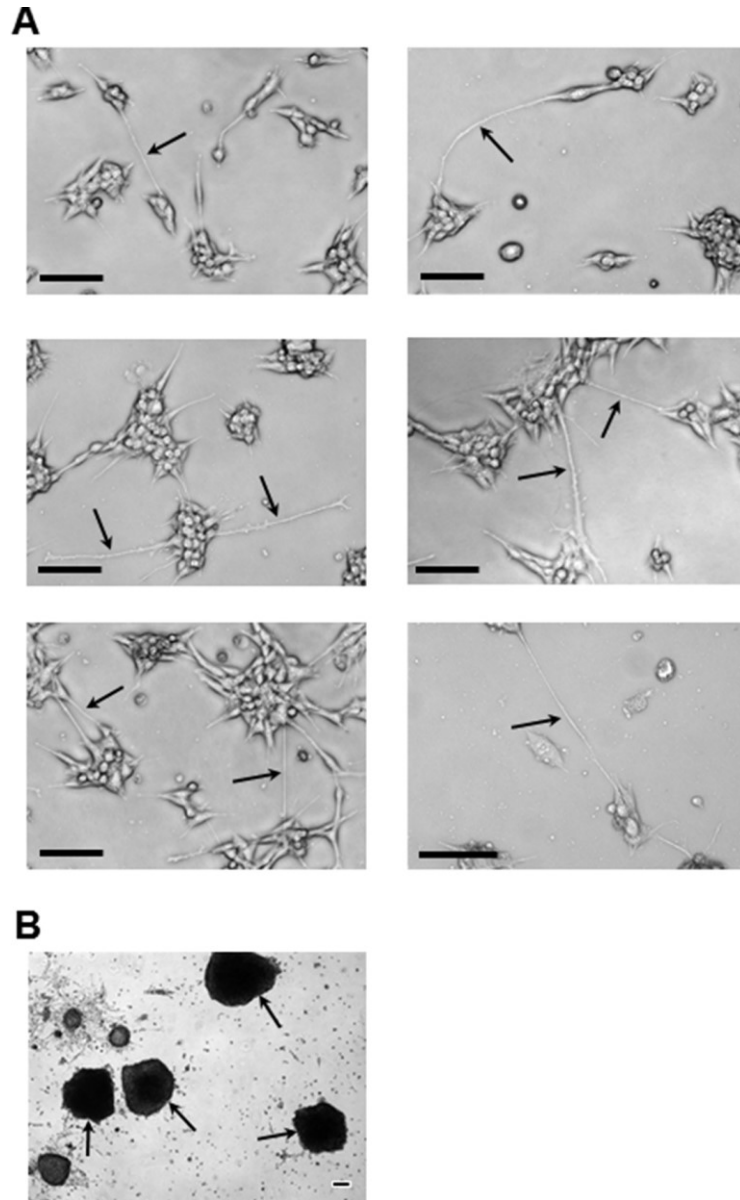
Figure 5.3 **Representative images of the brain cell line colony and mammosphere growth patterns.** (A) Images highlighting (arrows) the very long (> 100 μm) nanotube interconnections (or filopodia; e.g., middle right-hand image) that consistently form during: 24 hr (top row), 48 hr (middle row), and 120 hr (bottom row) of growth. (B) Examples, under adherent culture conditions, of the large free-floating mammospheres (arrows) that consistently formed during subculturing of smaller floating mammospheres retrieved from confluent brain cell line culture medium. Scale bars in all images depict 100 μm.

Overall, it was apparent that important distinctions between phenotypes and growth patterns (**Figure 5.3**) were present between cell lines. This and the very distinct non-

monolayer growth, ultimately reflect genetic diversity following metastasis and adaptation to the surrounding unique tissue environments.

Isogenic metastatic breast cancer cell lines: average specific growth rates. Average specific growth rates along with average cell cycle times (**Table 5.1**) were estimated by linear transformation[30] of the data used to generate the viable cell numbers versus days of growth curves presented in **Figure 5.4A**. Qualitative differences between the growth characteristics of the different cell lines can be ascertained from **Figure 5.4A**. Thus, it was apparent that primary tumor, liver, and spine cell lines appeared to have little or no lag-phase to their growth, followed by a relative steady rapid growth that ended with a plateau phase of slowed growth and the latter, for the liver cell line, occurred at a relatively low cell density. On the other hand, the brain cell line exhibited a protracted lag-phase that was followed by a rapid growth phase that did not reach a slowing of growth during the time period of this experiment. The lung cell line grew without a lag-phase, passed through a slowing of growth and then rapidly grew until the end of the experiment. The parental cell line had a relatively steady growth rate for most of the time that may have increased somewhat prior to the end of the experiment. The analysis of the plots shown in **Figure 5.4B** shed light on these qualitative evaluations. Thus, except for the parental cell line, all cell lines can be evaluated as having two distinct average specific growth rate periods shown as red tread lines in **Figure 5.4B**. The slopes of these lines provide estimates of average specific growth rates (μ) for each time period, which then allows for the calculation of the average length of the cell cycle (t_c) for each period (**Table 5.1**). As such, it was found that the parental cell line had an average specific growth rate of about 38% per day and a corresponding average cell cycle length of 1.8 days throughout the six days of growth. For

the brain, lung, and primary tumor cell lines the two periods of growth were from 24-72 hr and 72-144 hr (**Figure 5.4B**). While lung and primary tumor cell lines grew rapidly during the first time period, i.e., at average specific growth rates of 55% and 61% per day respectively, with corresponding short average cell cycle times of about 1.25 and 1.13 days respectively, the brain cell line exhibited a prolonged initial (24–72 hr) slow average specific growth rate of only about 15% per day that corresponded to an average cell cycle time of 4.65 days (**Table 5.1**). However, between 72 and 144 hr the brain cell line's average specific growth rate increased greater-than 3-fold to about 50% per day with an average cell cycle time of 1.4 days (**Table 5.1**). For the liver and spine cell lines the two periods of growth were from 24-72 hr and 96-144 hr (**Figure 5.4B**). These cell lines had similar growth characteristics with average specific growth rates of 43% and 46% per day respectively and corresponding average cell cycle lengths of 1.6 and 1.5 days respectively during the first time period (**Table 5.1**). Similarly, both liver and spine cell lines reached a stationary growth phase (96-144 hr) (**Figure 5.3B**) where average specific growth rates of only 7.6% and 6.2% per day and corresponding average cell cycle times that increased by nearly 10-fold to about 9 and 11 days respectively (**Table 5.1**).
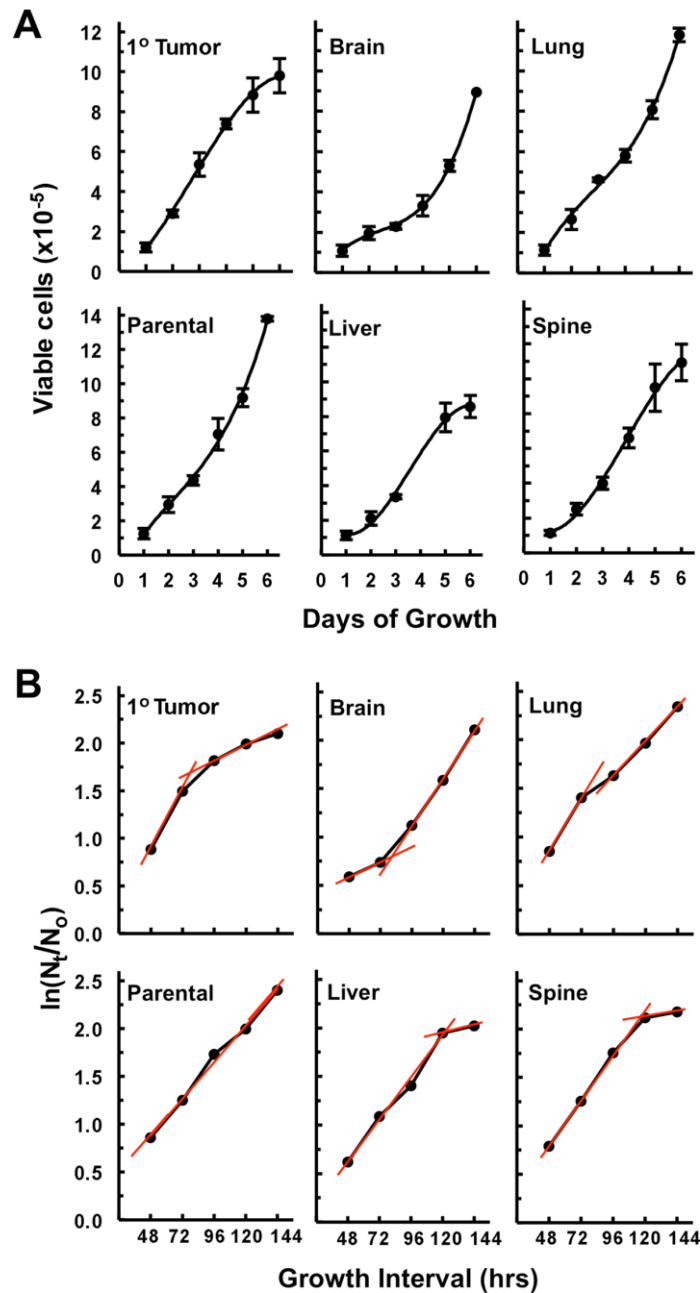
Figure 5.4 **Growth curves and estimation of average specific growth rates (μ) off of plots of ln(N$_t$/N$_o$) versus time.** (A) Growth curves of viable cell numbers vs. days of growth depicting distinctions in growth characteristics between cell lines. Each data point of the growth curves represents a mean (n = 3 to 4 wells of cells) ± 1 standard deviation except for the last point of the brain and the last two points of the liver where these are averages of two wells of cells. (B) The same data sets use in (A) plotted as ln(N$_t$/N$_o$) vs. growth interval in hr where N$_t$ is the number of cells at time 't', N$_o$ is the initial number of cells, i.e., viable cell counts on day 1 (24 hr after seeding the plates), and t is time. As, ln (N$_t$/N$_o$) = μt, it can be seen that the slope (μ) of each treadline (shown in red) provides an estimate of the average specific growth rates over the course of each growth interval (red lines) shown.

Overall, under these *in vitro* conditions, these results indicate that each of these isogenic cell lines modulates its cell cycle rate along with cell loss or senescence rate and hence its growth rate differently throughout the six-day culture period.

Isogenic metastatic breast cancer cell lines: motility. Two independent motility assays were carried out in standard 24 well Transwell® plates with 8 μm membrane inserts. During day two (white bars) and three (gray bars), the parental cell line's motility was relatively high as compared to the isolated isogenic cell lines (**Appendix Figure S5.2**). On day two, the motility of the parental cell line was significantly higher than the primary tumor and all metastatic cell lines ($P < 0.05$, two tailed *t*-test), and this remained the case on day three for all cell lines except the liver cell line. By day three the liver cell line's motility was significantly higher than the primary tumor and metastatic cell lines ($P < 0.005$) but not the parental cell line. In addition, it was noted that, except for the parental cell line, the numbers of cells migrating were very low being on average only 1.5% of the total cell numbers in the wells. The result is consistent with the fact that these metastatic cell lines do not exhibit extensive lateral monolayer growth patterns (**Figure 5.1-5.3**), which favors migration to and through a pore or into a "wound" but instead propagate vertically in stationary colonies.

Isogenic metastatic breast cancer cell lines: metabolomics. To gain a better understanding of the underlying molecular changes that are contributing to or are the result of adaption to different tissue microenvironments metabolomic analyses of our isogenic metastatic breast cancer cell lines were carried out. Our global metabolomics analyses provided strong evidence that our isogenic metastatic cell lines have distinct metabolomes (**Figure 5.5** and **Table 5.2**). Principal components analyses (PCA) of aqueous phase as

well as lipid phase (predominately lipids) metabolites (**Figure 5.5A-B**) revealed that all the isogenic cell lines unambiguously clustered into discrete classes (depicted as spheres) in both cases, which reflects each cell line's inherently distinct metabolome and lipid characteristics. PCA mapping of aqueous phase metabolites (**Figure 5.5A**) shows that PC 1 contributes to 49.9% and PC 2 contributes to 21.9% of the variation observed in the various samples. PCA demonstrates that there is a large difference, primarily indicated by PC 1, in metabolites of the metastatic cell lines relative to the primary tumor cell line (**Figure 5.5A**). Similarly, metastases are mainly differentiated by PC 2 indicating that they are more similar to one another. Hierarchical clustering confirmed that the primary tumor cell line was clustered separately from the metastases (**Figure 5.5A -** bottom panel). The dendrogram also confirmed that brain and liver metastases were more related to each other with respect to aqueous metabolite components. PCA mapping of lipid phase metabolites (**Figure 5.5B**) shows that PC 1 contributed to 45.3% and PC 2 contributed to 33.1% of the total variation. PCA showed that primary tumor lipids phase metabolites were closely related to brain lipid phase metabolites. Hierarchical clustering confirmed PCA analysis showing that lipid phase metabolites in primary tumor were closely related to brain (**Figure 5.5B -** bottom panel). The largest variation in lipid phase metabolites expression was observed in spine metastases.

Table 5.2 Fold increases of top metabolites from the 1° Tumor[1] and metastatic sites.

Metabolite	1° Tumor vs:			
	Brain	**Liver**	**Lung**	**Spine**
Biotripyrrin	4.2	14.0	7.5	34.6
2-Amino-3-Carboxymuconic Acid Semialdehyde	17.3	14.7	2.9	12.0
L-Thyroxine	----[2]	18.0	2.5	7.9
Phosphatidylinositol Trisphosphate (PIP3) (18:0/16:1)	3.5	4.6	6.7	3.9
Lysophosphatidylethanolamine (LysoPE) (15:0/0:0)	5.4	3.7	2.3	3.7
L-Dihydroorotic Acid	4.0	3.8	2.0	3.6
Cholesterol Ester (14:1)	2.7	3.3	4.8	3.7
Cholesterol Ester (20:4)	2.0	2.2	2.7	3.4
	Brain vs:			
	1° Tumor	**Liver**	**Lung**	**Spine**
Neurotensin 1-10	85	3.8	100	9.9
Tryptophyl-Tryptophan	60	24	----	36
Phosphotidylethanolamine (PE) (18:3/14:1)	23	2.2	----	9.6
Phosphotidylglycerolphosphate (PGP) (16:1/16:1)	8.0	4.5	----	----
	Liver vs:			
	1° Tumor	**Brain**	**Lung**	**Spine**
N1, N8 Diacetylspermidine	42	2.8	----	2.0
1-Phenylethylamine	2.5	2.1	19.7	5.2
Pantetheine	----	----	7.3	----
	Lung vs:			
	1° Tumor	**Brain**	**Liver**	**Spine**
CL[3]	----	----	----	----
Arginyl-Proline	4.6	5.3	22	3.4
DG[4] (14:0/24:1/0:0), (16:1/22:0/0:0), (18:1/20:0/0:0)	2.9	4.7	9.7	2.5
Putreanine	3.5	3.3	7.9	3.1
	Spine vs:			
	1° Tumor	**Brain**	**Liver**	**Lung**
Methionyl-Proline	71	20.8	3.1	44
Asparaginyl-Glutamate	----	26.6	16.6	----

Gentisate Aldehyde	----	26.6	16.6	----
Dityrosine	22.4	12.6	2.5	4.8
PIP2 (16:0/20:1), (16:0/22:4) & (18:0/18:1)	21.8	15.3	2.3	20.5
4-Guanidinobutanoic Acid	----	20.9	----	2.3
5-Methyldeoxycytidine	----	16.6	4.5	----
Ferrocytochrome	6.4	14.8	8.9	6.9
Pentacaboxylporphrinyl	----	10.2	5.2	5.5
SCICAR[5]	----	7.9	4.3	8.0
4-Aminobutyraldehyde	6.1	6.1	5.3	2.5
N2, N2-Dimethylguanosine	----	6.6	6.9	3.3
D-Lactaldehyde	6.2	6.6	6.1	2.5
7,8-Dihyderoneopterin	3.1	3.9	2.9	2.0

[1]1° Tumor denotes primary tumor.

[2]A dash indicates that the metabolite was below the detection limit in that organ.

[3]CL denotes cardiolipins: (16:0/16:0/18:1(9Z)/18:1(9Z)), (16:0/16:0/18:1(11Z)/18:1(11Z)), (16:0/16:0/18:1(9Z)/18:1(11Z)), & (16:1 (9Z)/18:0/16:1(9Z)/18:0).

[4]DG denotes diacylglycerol.

[5]SCIAR denotes: (S)-2-[5-Amino-1-(5-phospho-D-ribosyl) imidazole-4-carboxamido]succinate.
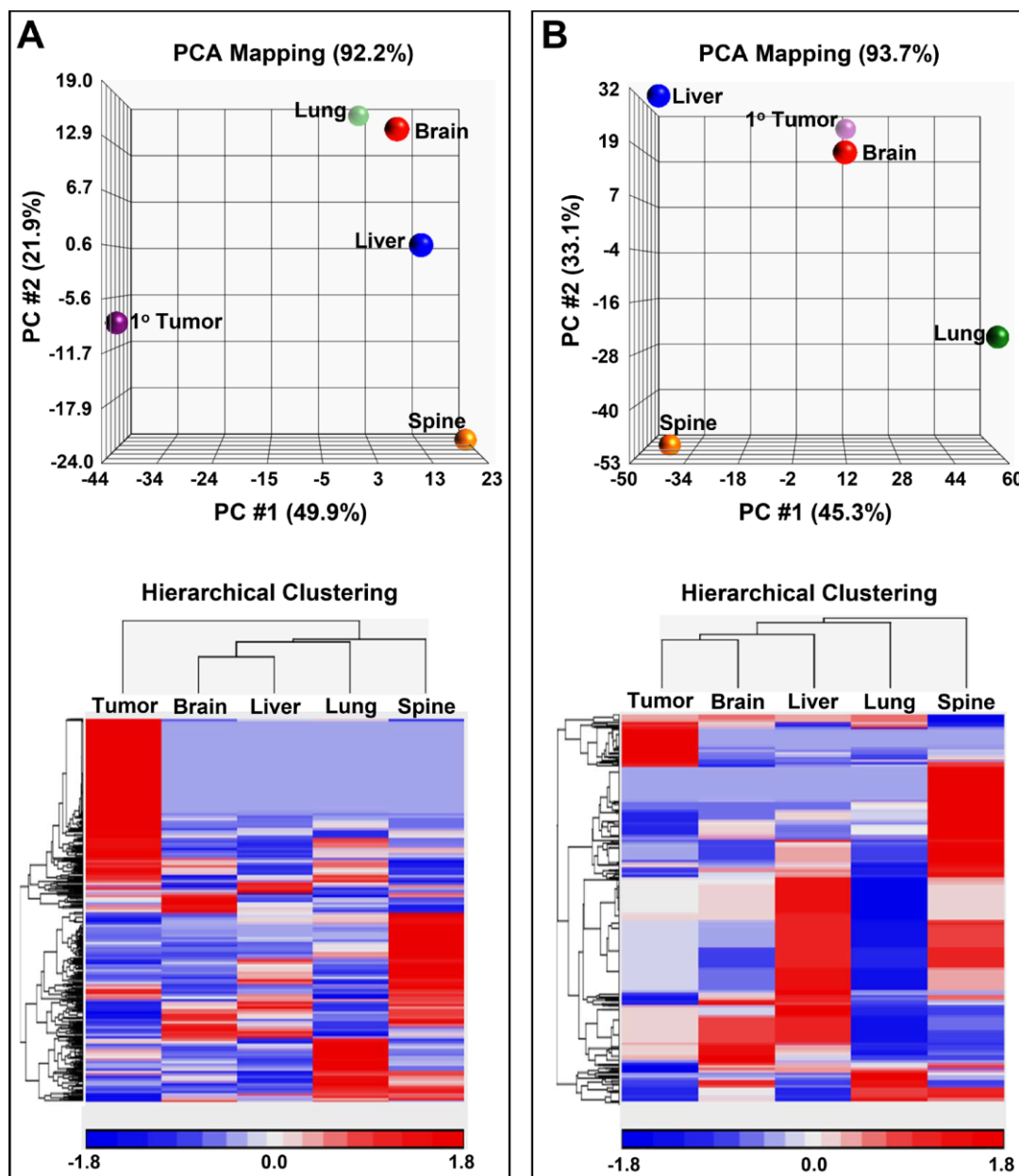
Figure 5.5 **Principal component analysis (PCA) maps along with hierarchical clustering's of metabolites and lipids.** (A) 3D PCA mapping of aqueous metabolites (top panel) displaying sample classes as spheres. Bottom panel displays hierarchical clustering of the samples along with the associated heat map of aqueous metabolite distributions. (B) 3D PCA mapping of lipid soluble metabolites (top panel) with spheres representing the sample classes. Panel at the bottom displays a heat map of lipid soluble metabolite distributions along with the associated dendrogram. Expression values for the heat maps are indicated by a key at the bottom of the maps.

Raman spectroscopic differentiation of organ-specific metastatic isogenic breast cancer cell lines. Mean Raman spectra (±1 standard deviation) of the IMBC cell lines and the primary tumor cell line are shown in **Figure 5.6A**, where the spectra have been normalized and offset for visualization purposes but displayed without background correction. The Raman spectra are an aggregate expression of cellular biochemistry and structure, since the vibrational signatures inform not only on the composition of the complex biological material but also on structural states of the molecules in the specimen. The observed spectral features encode a vast amount of information of the principal constituents[43], namely lipids, proteins, nucleic acids, carbohydrates and small molecules. **Table 5.3** lists the assigned vibrational modes for a selection of these spectral features. Though the spectra grossly appear to have similar profiles, detailed inspection reveals subtle but discernible and reproducible shape differences, especially on removal of the broad fluorescence background[44]. Based on our previous experience in differentiation of breast tissue lesions[21], we reasoned that while the subtle distinctions between the spectra of each cell line impede the possibility of differentiation using a single feature, multivariate classification methods could enable recognition and segmentation of the cell pathology - as long as the between-class distinctions are reproducible and surpass within-class distinctions.

Table 5.3 Assignment of specific Raman spectral components to subcellular constituents.

Raman Shift (cm⁻¹)	DNA/RNA	Proteins	Lipids
620		C-C Twist Aromatic Ring	
643		C-C Twisting of Tyrosine	
702			Cholesterol
715	Adenine		C-N (Membrane Phospholipid Head Group), $CN^-(CH_3)_3$
781	Cytosine, Uracil Ring Breathing		
810			Phosphodiester
828	O-P-O Stretching	Tyrosine	Phosphodiester
853		Ring Breathing of Tyrosine, C-C Stretch of Proline Ring	
878			$C-C-N^+$ Symmetric Stretching
938		Hydroxyproline, Proline, ν(C-C) Vibration of Collagen Backbone	
1005		Phenylalanine	
1035		Collagen	
1066		Proline	Fatty Acid
1086			$\nu_1CO_3^{2-}$, $\nu_3PO_4^{3-}$, ν(C-C)
1128		C-N Stretching	
1156		C-C, C-N Stretching	
1176		C-H Bending Tyrosine	
1209		Tryptophan, Phenylalanine	
1241			Asymmetric Phosphate Stretching
1254		C-N in Plane Stretching	
1266		Amide III (α-Helix), Tryptophan, Collagen	
1302		$CH_3/\delta(CH_2)$ Twisting, Wagging, Collagen, Amide III & Methylene bending mode	$CH_3/\delta(CH_2)$ Twisting, Wagging, Phospholipids & Methylene bending mode
1334	Nucleic Acid	CH_3CH_2 Wagging, Collagen	

1342		CH Deformation	
1367			$v(CH_3)$, Phospholipids
1391		C-N Stretching in Quinoid, Ring-Benzenoid	
1437			CH_2 Deformation
1451		CH_2 Bending, CH_3 Bending, C-H Deformations	C-H Deformation
1556		Tryptophan $v(CN)$, Amide II $v(C=C)$ Porphyrin & Tyrosine	
1605	Cytosine	Phenylalanine & Tyrosine	
1657		$v(C=O)$ Amide I (α-Helix) C=O Stretching of Collagen & Elastin	C=C Stretch, Fatty Acids

v, stretching mode; δ, bending mode.

Figure 5.6 **Raman spectroscopic analyses of organ-specific metastatic breast cancer cell lines reveals distinct spectral characteristics for each cell line.** (A) Representative Raman spectra acquired from brain, primary tumor (1° Tumor), liver, lung, and spine cell lines. The solid profile depicts the mean spectrum of each sample group and the shadow represents ±1 standard deviation. Spectra were normalized and offset for visualization. Dashed vertical lines delineate Raman shifts (cm⁻¹) detailed in **Table 5.3**. (B) Principal component (PC) loadings for PC 1, 2, 3 and 5, for the Raman measurements are shown. Dashed vertical lines delineate prominent Raman shifts (cm⁻¹) detailed in **Table 5.3**. (C) Radial visualization principal component scores plot, corresponding to the most discriminative PCs (PC 1, 2, 3 and 5), shows the clustering of the spectral data corresponding to each organ-specific cell line, red: primary tumor, blue: brain, green: liver, orange: lung, and purple: spine. (D) Dendrogram of organ-specific breast cancer cell lines cluster analysis. Each color bar represents one organ-specific cell line. (E) Identification of informative spectral regions via PCA data exploration as exemplified by the PC loadings

corresponding to the spectral dataset acquired from: primary tumor and liver (left panel) and primary tumor and spine (right panel) cell lines. The top to bottom profiles in each panel show difference spectra: (DS) between liver/primary or spine/primary spectra along with their PC 1 and PC 2 loadings, respectively. The highlighted yellow bars (1–4), represent the wavelength regions elucidated from the difference spectra (DS) as those with the most significant variability amongst the considered cell lines.

Hence, we employed principal component analysis (PCA) to transform the dimensions of the acquired spectra into a set of linearly uncorrelated variables, i.e., principal components, along which the variation in the data is maximal (**Figure 5.6B**). 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. PC 1 and PC 2 accounts for approximately 67% and 12% of the total variance in the dataset. In addition to homing in on the spectral features responsible for the variance between the cell lines (indicated by the dashed lines in **Figure 5.6B**), we employed the PC scores to assess the feasibility of recognizing individual cell lines based on the Raman data. Specifically, the PC scores were used to create a radial visualization plot (**Figure 5.6C**). The plot reveals the degree of clustering of the spectra recorded from the same cell line and, critically, the inter-cell line spectral variations. Together, these qualitatively suggest the presence of differential molecular constituents in the isogenic cell lines that are driven by site-specific adaptations. The Raman spectra-derived dendrogram (**Figure 5.6D**), reinforces these feasibility results but also hints at the relative difficulty in separating the liver and brain cell lines based solely on the vibrational signatures. Notably, the overlap between the brain and liver Raman signatures is consistent with the metabolomics findings (**Figure 5.5**) underscoring the correspondence between the two complementary data sets.

To quantify the classification capability of Raman spectroscopy, we developed decision algorithms based on partial least squares discriminant analysis (PLS-DA) and support vector machines (SVM). 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% (**Table 5.4**). The SVM-derived decision algorithm also provides similar levels of classification performance affirming that the richness of the spectral data is the principal driver for the prediction performance. Next, we performed difference analyses across the normalized spectra obtained from pairwise comparison of cell lines to delineate the informative regions with the goal of identifying biomarkers, which would be either universal or characteristic to a specific pair of cell lines. **Figure 5.6E** exhibits two representative cases of these comparisons, namely between primary tumor and liver cell lines and between primary tumor and spine cell lines. The accompanying PC loadings were obtained from analysis of the spectral dataset constituted by the primary tumor and liver, and primary tumor and spine cell lines, respectively. By merging the difference analyses and spectroscopic basis of the PC loadings, the following informative regions were identified: $1000–1006$ cm^{-1}, $1136–1211$ cm^{-1}, $1298–1330$ cm^{-1} and $1435–1470$ cm^{-1}.

Table 5.4 Raman spectroscopy-based classification of IMBC cell lines.

Reference Identification	PLSDA[1] algorithm Correct Classification (%)	SVM[2] algorithm Correct Classification (%)	PLSDA[1] feature-specific algorithm Correct Classification (%)
1° Tumor[3]	99.3 (0.7)[4]	98.9 (1.1)	97.2 (2.8)
Brain	98.0 (2.0)	99.6 (0.4)	91.7 (8.3)
Liver	97.4 (2.6)	94.3 (5.7)	91.1 (8.9)
Lung	93.3 (6.7)	97.3 (2.7)	85.8 (14.2)
Spine	96.1 (3.9)	98.2 (1.9)	90.6 (9.4)

[1]Denotes Partial Least Squares Discriminant Analysis.
[2]Denotes Support Vector Machines.
[3]Denotes Primary Tumor cell line.
[4]Values in parenthesis = percentage misclassifications.

Using only the selected regions (highlighted by the yellow bars of **Figure 5.6E**), we developed a PLSDA-derived decision algorithm to reclassify all the cell lines that provided equally impressive prediction performance (**Table 5.4**) as that obtained using the full spectral analysis. Only 9.6% of the spectral information was used in this case thereby underlining the presence of specific spectral markers in the dataset.

Additionally, to ensure the robustness of these findings, we 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, we 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%. This underscores the robustness of the spectroscopic measurements to confounding variables and chance correlations. Collectively, these results demonstrate that Raman spectroscopy offers a reliable tool for

discriminating these IMBC cell lines on the basis of distinct organ-of-origin driven biochemical adaptations.

We also compared Raman signatures of specific cell line pairs and tallied the spectral markers against the known Raman features of the cell line specific expressed metabolites. As show in in **Figure 5.7** - top panel, a comparison of metabolite profiles in the spine and the primary cell line reveals the overexpression of Raman-active analytes in spine, namely gentisate aldehyde and dityrosine. Similarly, complementarity of the metabolomics and Raman datasets was reinforced through detection (using liver as a control) of the overexpressed Raman-active analytes in the primary tumor: L-thyroxine and L-dihydroorotic acid (**Figure 5.7** - middle panel) and Raman-active 1-phenylethylamine in the liver (using primary tumor as a control; (**Figure 5.7** - bottom panel). Analyses of the difference spectra between cell line pairs and their corresponding PC loadings reveals the presence of subtle features at wavenumbers (scattering frequencies) where these metabolites show Raman activity (**Figure 5.7** and **Table 5.5**).
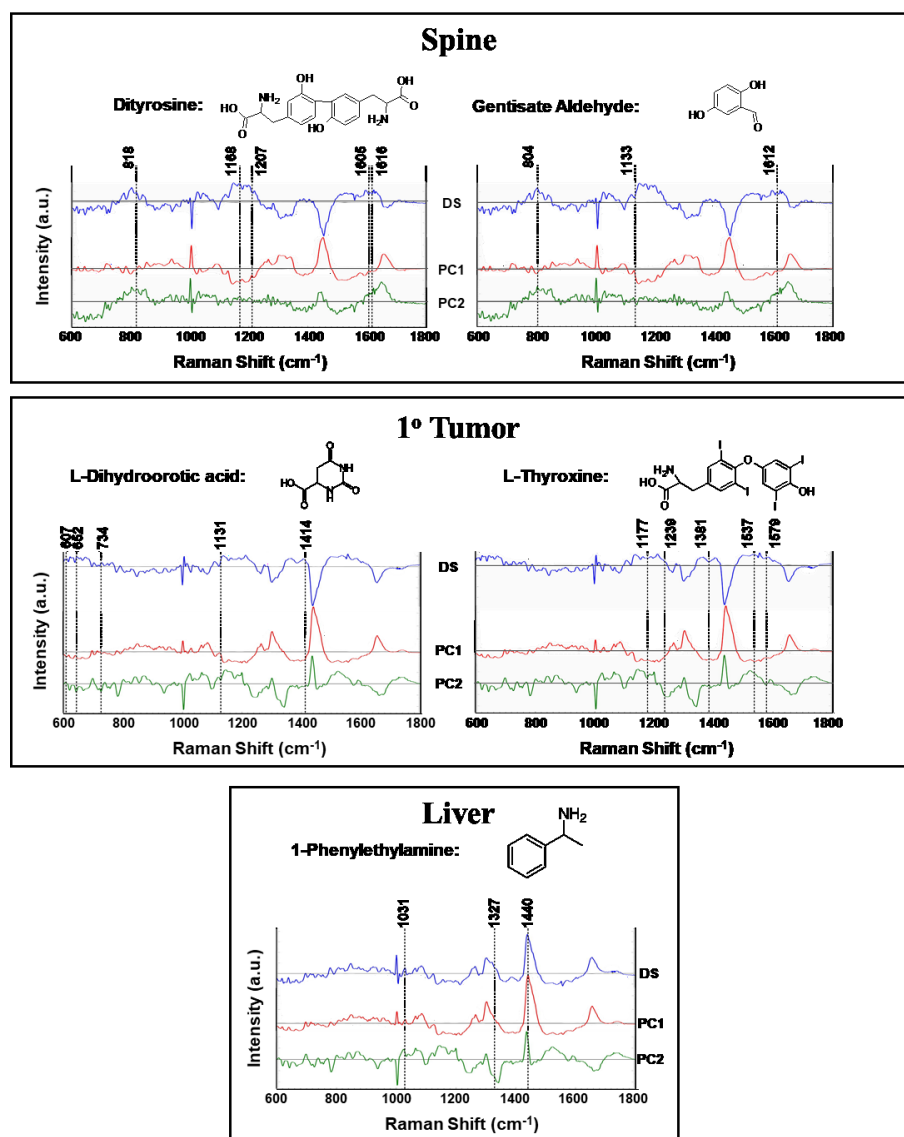
Figure 5.7 **Raman spectroscopic analysis to probe the presence of cell line specific expression of molecules identified through metabolomics analysis.** The top panel highlights the differential expression of spectral markers in the spine cell line. The primary cell line spectrum was used as the control to calculate the difference profiles. Additionally, principal components (PC) 1 and 2, calculated from the spine and primary cell line data, are provided to capture the variance. The presence of spectral features, corresponding to the peaks of dityrosine and gentisate aldehyde, are highlighted by the dashed lines and detailed vibrational mode assignment is presented in **Table 5.5**. Similarly, the middle panel compares the Raman spectra of the primary cell line with a control group, i.e., Raman spectra acquired from the liver cell line, to illustrate the presence of features of overexpressed metabolites L-dihydroorotic acid and L-thyroxine. The bottom panel compares the Raman profiles of the liver cell line with the control group (primary) to delineate the overlap with features of 1-phenylethylamine. Profiles in blue represent the difference spectra: (DS) whereas the red and green profiles show the PC 1 and PC 2 loadings respectively for each chosen pair of cell lines.

Table 5.5 Raman shifts and associated band assignments of cell line specific metabolites.

Tissue	Metabolite	Raman Shift (cm^{-1})	Band Assignment
Spine	**Dityrosine**	818	Tyrosine, Proline, Hydroxyproline, v_2PO^{2-} stretch of nucleic acids
		1168	Tyrosine: C-H in-plane bending
		1207	Tyrosine & Phenylalanine: $C-C_6H_5$ stretching, Hydroxyproline
		1605	Tyrosine & Phenylalanine: C=C in-plane bending
		1616	Tyrosine & Tryptophan: C=C stretching mode
	Gentisate Aldehyde	804	vC-C, vC-O
		1133	b C-H, vC-C
		1612	vC-C, bC-H
1° Tumor	**L-Dihydroorotic Acid**	607	βC=O
		652	γNH, γOH
		734	βC=O, vC=O
		1131	vNC, βCH
		1414	vNC, βNH
	L-Thyroxine	1177	Out-of-phase vC=O
		1239	In-phase vC=O
		1381	Stretching aromatic ring
		1537	In-phase aromatic ring
		1579	Stretching aromatic ring
Liver	**1-Phenylethylamine**	1031	C-H in-plane bending mode, C-N stretching
		1327	CH_3 wagging mode
		1440	CH_3, CH deformation vibrations

1° Tumor denotes Primary Tumor.
b, bending; v, stretching; β, in-plane deformation; γ, out of plane.

5.5 Discussion

An inherent challenge within cancer research is the cataloguing of fundamental information on what is generally fatal metastatic disease within vital organs. Monitoring and treating metastatic progression remains a formidable task due to many gaps in our knowledge including: an inability to monitor the onset of successful metastatic growth along with specific differential molecular adaptations that allow for the cancer to survive and thrive within different tissue types[18,45,46]. Consequently, we have taken up the important consideration that metastatic cancer cells growing in visceral organs ought not to be considered simply as primary tumor implants. Rather metastatic lesions need to be understood as significantly influenced and altered by tissue-specific microenvironments that the cancer cells must adapt to[18,45,46]. To address this problem, we choose to characterize IMBC cell lines that spontaneously arose from dissemination from the primary mammary fat pad site of our mouse model.

The establishment of metastatic lesions integrated into vital visceral organs is a multistep process that includes: i) the cancer cells ability to survive as independent entities that thrive outside the normal cell-cell interactions of healthy epithelial tissues, ii) surviving harsh circulation conditions, and iii) embed and adapt to growth within microenvironments that are distinct, from a developmental as well as functional basis, from the primary tumor site. In order to mimic this process, we have generated isogenic cell lines directly from organ-specific metastatic lesions of the brain, liver, lung, and spine (bone), which are the organs most commonly affected, i.e., bone (60%), lung (34%), liver (20%), and brain (10-15%), during human metastatic breast cancer progression[25,26]. We initially cultured these as organ explants along with a cell line from the primary tumor site, i.e., mammary fat pad,

in an attempt to preserve tissue adaptation attributes acquired during adaptation to each *in vivo* microenvironment[47,48]. In addition, these culturing conditions minimized damage as well as stresses imposed during harsher multistep single cell isolation protocols, which likely also bias cancer cell selection to subpopulations that survive the isolation procedures. This overall model is analogous to the natural course of metastatic progression found in the clinic where metastatic lesions are composed of cells that are isogenic to the primary breast cancer but also distinct at the molecular and cellular level[6-18,45,46].

Once adapted to plastic these new isogenic cell lines exhibited distinct phenotypic/morphological differences in growth patterns along with similarities that went across cell lines including those established from the primary tumor site. The most evident of the latter was a tendency for all cell lines to grow as complex arrays of interconnected 3D colonies with various degrees of loosely held together monolayers (**Figure 5.1-5.3**). The distinctions in average specific growth rates and average cell cycle rates (**Figure 5.4** and **Table 5.1**) support the concept that the different growth patterns reflect metabolic, cell cycle, and hence, likely genetic/epigenetic differences between cell lines. Given that, in general, cytotoxic chemotherapies are more efficient in killing cycling cells, understanding differences in growth and cell cycle rates of cells in metastatic lesion will provide us with an optimum treatment strategy[49].

Attempts to distinguish between cell lines using motility assays proved inconclusive (**Appendix Figure S5.2**) possibly due in part to the relative lack of monolayer growth exhibited to different degrees across cell lines (**Figure 5.1-5.3**), which is the type of lateral growth pattern that generally propagates a migration to and through a pore or into a "wound". It is also possible that the low motility reflects a loss of this metastatic trait that

is generally associated with movement out of the primary tumor, into the blood/lymph systems, and subsequent local/systemic dissemination, which may be suppressed once growth is established in a distal organ. Nevertheless, the little or no motility seen across all cell lines in the Transwell motility assay (**Appendix Figure S5.2**) revealed that these cell lines were able to survive and grow in the low nutrient (0.1% FBS) serum conditions of the top chamber, which may reflect an adaptation to survival in a hostile microenvironment within portions of the primary tumor prior to dissemination or at the new site of growth. From a different perspective, these cell lines are 'motile' in up-ward growth in the form of 3D spheres and all exhibit a tendency to shed viable free-floating cells into the medium that either remained as single cells or grew as free-floating colonies that resemble mammospheres (**Figure 5.3B**). The latter trait is unusual as mammosphere growth patterns have consistently been shown to be limited to growth on non-adherent plates[50] unlike the adherent conditions used here. This trait may reflect an *in vivo* attribute of the metastatic process or adaptation to plastic. The latter seems unlikely as most available cancer cell lines adapted to growth on adherent plastic do so as monolayers with little or no "sphere" formations.

Interestingly, the colonies of brain and spine (and to a lesser extent the other cell lines as well) are interconnected and therefore, in apparent communication[51-53], by nano- or microtubes, which were observed at well over 100 μm in length (**Figure 5.2** and **5.3**) and at high colony densities formed complicated intricate networks between colonies (**Figure 5.2E-H**). It appears that under these conditions the cell lines have a propensity to grow as semi-separate entities/colony arrays that require an interacting exchange of materials via these conduits[51-53]. This may be a reason for the lag-phase growth period exhibited by the

metastatic brain cell line as these cells may require a relatively extensive interconnected network to support higher growth rates and shortened cell cycle times. To the best of our knowledge, at the abundance seen in our cultures, the interconnected nanotube network is a very unique characteristic along with the mammosphere formation on adherent plates.

Importantly, the metabolomics data indicates that the isogenic metastatic cell lines have diverged from the primary tumor as well as from each other (**Figure 5.5**). **Table 5.2** shows fold increases of top metabolites, i.e., those at least 2-fold greater in each tissue-specific cell line vs all other cell lines. Although further work is needed to definitively prove whether an increase in a metabolite in a specific metastatic cell line has arisen from organ-of-origin influences, select metabolites in **Table 5.2** can be associated with specific organs. For example, as reflected in **Table 5.2**, Neurotensin 1-10, a neurotransmitter, was found to be increased in the brain cell line and has been reported to be principally of brain origin[54] while pantetheine (vitamin B5) an intermediate in the enzyme-CoA formation pathway, which is increased in the liver cell line, is generally most abundant in liver[55]. Some classes of the tetra-acylated anionic phospholipids: cardiolipins, are only found in relatively high abundance in the lung cell line (**Table 5.2**) and cardolipins have been reported to be increased in human lung cancer[56]. Interestingly, N1, N8 diacetylspermidine has been found to be a marker of breast cancer and from our results (**Table 5.2**) it appears that it can potentially reflect metastatic progression to the liver[57]. Future studies are required to provide experimental evidence that cell line specific metabolomes contain metabolites that reflect a cell line's tissue of origin. Such work would also strive to obtain information on metabolic pathways associated with such metabolites and their potential relationships to metastatic adaptations at each site.

We also had the aim of obtaining rapid, non-destructive, and label-free profiling of these IMBC cell lines and to this end have undertaken a Raman spectroscopy characterization approach (**Figure 5.6**). Raman spectroscopy was considered as a complementary alternative to a purely "omics" approach as the latter has some well-characterized limitations[33]. Our Raman spectroscopy-based decision algorithms showed the ability to differentiate between our isogenic cell lines with high accuracy (**Table 5.4**). 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. The complementary nature of these distinct analytical tools (metabolomics and Raman spectroscopy) was observed; e.g., with the general overlap that was found between the dendrograms obtained from the two methods, which in both cases indicated that brain and liver cell lines are closely related.

Notably, we sought more evidence of a complementarity between the two methods and found examples of Raman-active analytes, i.e., discriminating spectral markers were ascertained for cell line specific metabolites (**Figure 5.7**). As discussed earlier, the principal variations in the Raman spectra of the cell lines are largely attributable to proteins, lipids and nucleic acids however, signatures of metabolites and other small molecules are also embedded in the Raman spectra. Hence, even though fingerprinting specific metabolites through the vibrational features alone is challenging, one can infer the contributions of these metabolites towards the composite cellular biochemical status represented in the Raman data. Further probing of the high wavenumber region may provide complementary molecular insights, particularly of the lipid phenotype along with other important biochemical features[58-62].

Overall, important differences between organ-specific metastatic cell lines reflect the fact that organs differ vastly with unique attributes of metabolism, developmental programs, microenvironments, and function, all of which results in defined identities with specific growth challenges for invading cancer cells. For example, normal oxygen tension varies greatly between tissues[63]. Therefore, if one considers only this single vital nutrient change between organ types, it ought not to be surprising that a metastatic growth embedded in lung tissue with high oxygen tension would acquire different characteristics as compared to metastatic cells thriving in brain or bone at a much lower oxygen levels[63,64].

5.6 Appendix
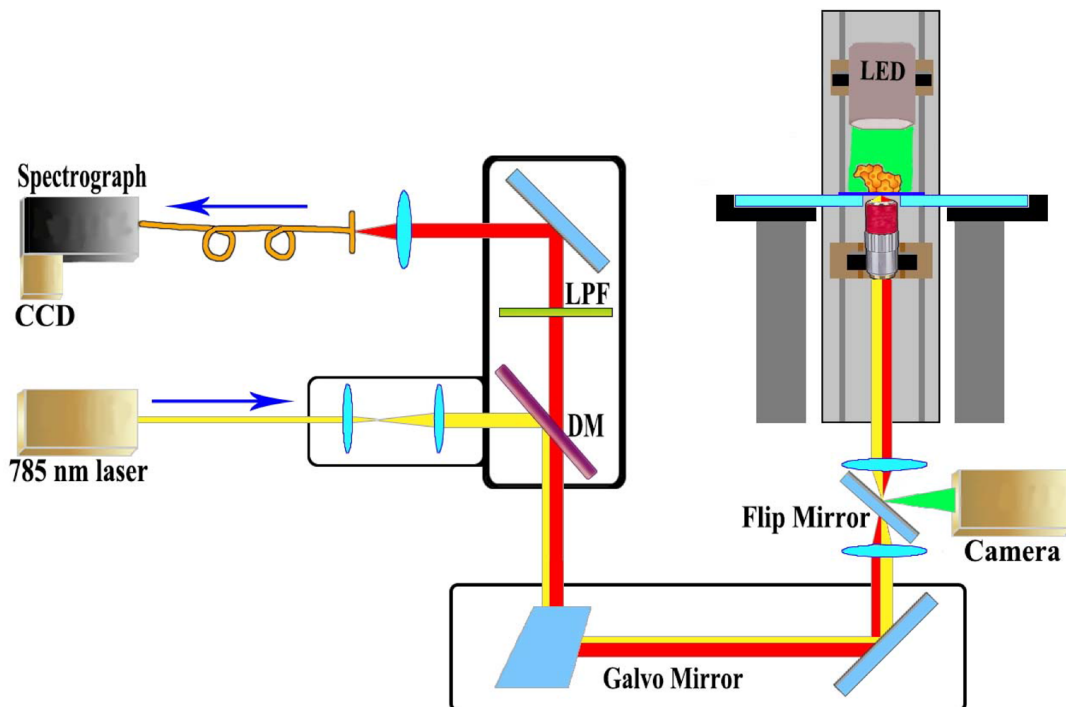
Figure S5.1 **Schematic illustration of Raman microspectroscopy system.** The system incorporates confocal Raman, confocal reflectance (not shown here) and bright field imaging modalities for visualization and characterization of unstained live cells. LPF: Long Pass Filter; DM: Dichroic Mirror.
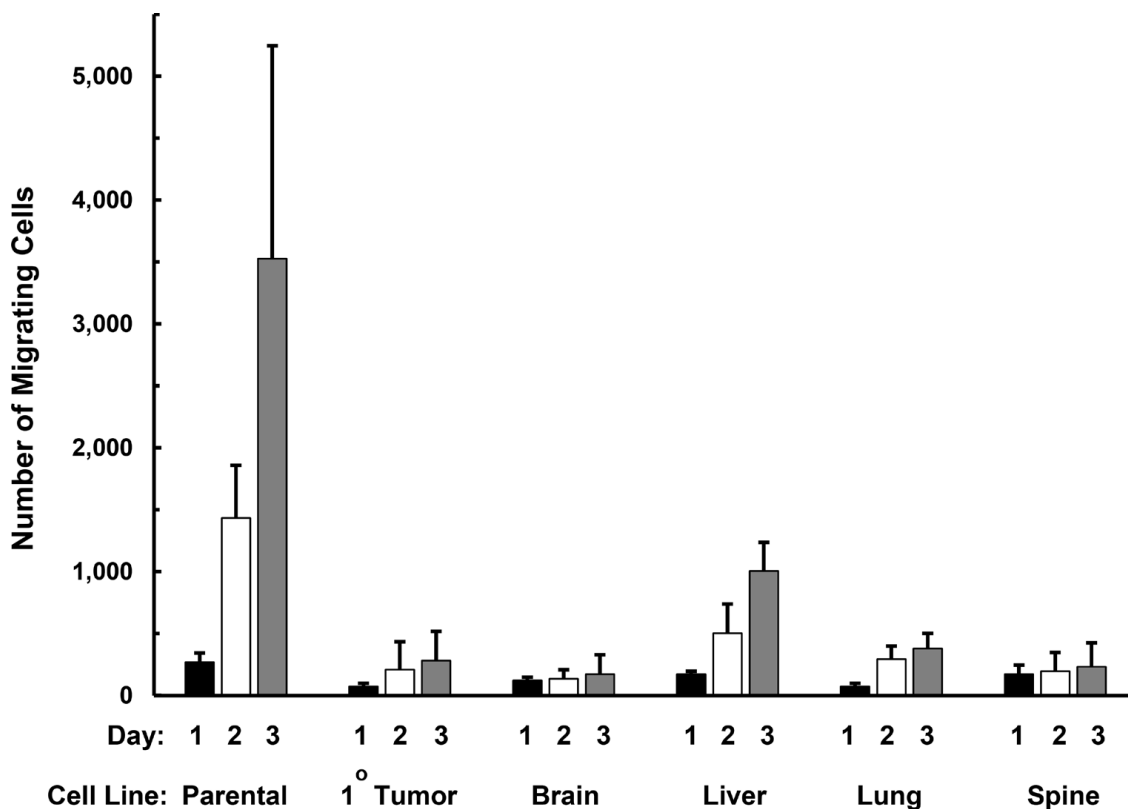
Figure S5.2 **Evaluation of motility through porous (8 µm) membrane inserts in standard 24 well Transwell® plates.** During day two (green bars) and three (red bars), the parental cell line's motility was relatively high as compared to the isolated isogenic cell lines. On day two, the motility of the parental cell line was significantly higher than the primary tumor and all metastatic cell lines ($P < 0.05$, two tailed t-test) and this remained the case on day three for all cell lines except the liver cell line. By day three the liver cell line's motility was significantly higher than the primary tumor and metastatic cell lines ($P < 0.005$) but not the parental cell line. Error bars depict ± 1 SD.

Acknowledgements

Bibliography

1. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2011 Incidence and Mortality Web-based Report. **2014**.

2. DeSantis, C.E., Lin, C.C., Mariotto, A.B., Siegel, R.L., Stein, K.D., Kramer, J.L., Alteri, R., Robbins, A.S. and Jemal, A. Cancer treatment and survivorship statistics, 2014. *CA. Cancer J. Clin.* **2014**, 64:252–271.

3. Nola, S., Sin, S., Bonin, F., Lidereau, R. and Driouch, K. A methodological approach to unravel organ-specific breast cancer metastasis. *J. Mammary Gland Biol. Neoplasia.* **2012**, 17:135–145.

4. Jacot, W., Gutowski, M., Azria, D. and Romieu, G. Adjuvant early breast cancer systemic therapies according to daily used technologies. *Crit. Rev. Oncol. Hematol.* **2012**, 82:361–369.

5. Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., Abbott, R.M., Hoog, J., Dooling, D.J., et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* **2010**, 464:999–1005.

6. Amir, E., Miller, N., Geddie, W., Freedman, O., Kassam, F., Simmons, C., Oldfield, M., Dranitsaris, G., Tomlinson, G., Laupacis, A., Tannock, I.F. and Clemons, M. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *J. Clin. Oncol*. **2012**, 30:587–592.

7. Aurilio, G., Disalvatore, D., Pruneri, G., Bagnardi, V., Viale, G., Curigliano, G., Adamoli, L., Munzone, E., Sciandivasci, A., De Vita, F., Goldhirsch, A. and Nole, F. 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.* **2014**, 50:277–289.

8. Aurilio, G., Monfardini, L., Rizzo, S., Sciandivasci, A., Preda, L., Bagnardi, V., Disalvatore, D., Pruneri, G., Munzone, E., Della Vigna, P., Renne, G., Bellomi, M., Curigliano, G., et al. Discordant hormone receptor and human epidermal growth factor receptor 2 status in bone metastases compared to primary breast cancer. *Acta. Oncol.* **2013**, 52:1649–1656.

9. Botteri, E., Disalvatore, D., Curigliano, G., Brollo, J., Bagnardi, V., Viale, G., Orsi, F., Goldhirsch, A. and Rotmensz, N. Biopsy of liver metastasis for women with breast cancer: impact on survival. *Breast*. **2012**, 21:284–288.

10. Criscitiello, C., Andre, F., Thompson, A.M., De Laurentiis, M., Esposito, A., Gelao, L., Fumagalli, L., Locatelli, M., Minchella, I., Orsi, F., Goldhirsch, A. and Curigliano, G. Biopsy confirmation of metastatic sites in breast cancer patients: clinical impact and future perspectives. *Breast Cancer Res*. **2014**, 16:205.

11. de Duenas, E.M., Hernandez, A.L., Zotano, A.G., Carrion, R.M., Lopez-Muniz, J.I., Novoa, S.A., Rodriguez, A.L., Fidalgo, J.A., Lozano, J.F., Gasion, O.B., Carrascal, E.C., Capilla, A.H., Lopez-Barajas, I.B., et al. Prospective evaluation of the conversion rate in the receptor status between primary breast cancer and metastasis: results from the GEICAM 2009–03 ConvertHER study. *Breast Cancer Res. Treat*. **2014**, 143:507–515.

12. Dieci, M.V., Barbieri, E., Piacentini, F., Ficarra, G., Bettelli, S., Dominici, M., Conte, P.F. and Guarneri, V. Discordance in receptor status between primary and recurrent breast cancer has a prognostic impact: a single-institution analysis. *Ann. Oncol.* **2013**, 24:101–108.

13. Guarneri, V., Giovannelli, S., Ficarra, G., Bettelli, S., Maiorana, A., Piacentini, F., Barbieri, E., Dieci, M.V., D'Amico, R., Jovic, G. and Conte, P. Comparison of HER-2 and hormone receptor expression in primary breast cancers and asynchronous paired metastases: impact on patient management. *Oncologist*. **2008**, 13:838–844.

14. Klein-Goldberg, A., Maman, S., Witz, I.P. The role played by the microenvironment in site-specific metastasis. *Cancer Lett.* **2014**, 352:54–58.

15. Liedtke, C., Broglio, K., Moulder, S., Hsu, L., Kau, S.W., Symmans, W.F., Albarracin, C., Meric-Bernstam, F., Woodward, W., Theriault, R.L., Kiesel, L., Hortobagyi, G.N., Pusztai, L., et al. Prognostic impact of discordance between triple-receptor measurements in primary and recurrent breast cancer. *Ann. Oncol.* **2009**; 20:1953–1958.

16. Santinelli, A., Pisa, E., Stramazzotti, D. and Fabris, G. HER-2 status discrepancy between primary breast cancer and metastatic sites. Impact on target therapy. *Int. J. Cancer.* **2008**, 122:999–1004.

17. Thompson, A.M., Jordan, L.B., Quinlan, P., Anderson, E., Skene, A., Dewar, J.A., Purdie, C.A. and Breast Recurrence in Tissues Study G. Prospective comparison of switches in biomarker status between primary and recurrent breast cancer: the Breast Recurrence In Tissues Study (BRITS). *Breast Cancer Res.* **2010**, 12:R92.

18. Wikman, H., Lamszus, K., Detels, N., Uslar, L., Wrage, M., Benner, C., Hohensee, I., Ylstra, B., Eylmann, K., Zapatka, M., Sauter, G., Kemming, D., Glatzel, M., et al. Relevance of PTEN loss in brain metastasis formation in breast cancer patients. *Breast Cancer Res.* **2012**, 14:R49.

19. Pinder, S.E., Rakha, E.A., Purdie, C.A., Bartlett, J.M., Francis, A., Stein, R.C., Thompson, A.M., Shaaban, A.M. and Translational Subgroup of the NBCSG.

Macroscopic handling and reporting of breast cancer specimens pre- and post-neoadjuvant chemotherapy treatment: review of pathological issues and suggested approaches. *Histopathology.* **2015**, 67:279–293.

20. Rehman, S., Movasaghi, Z., Tucker, A.T., Joel, S.P., Darr, J.A., Ruban, A.V. and Rehman, I.U. Raman spectroscopic analysis of breast cancer tissues: identifying differences between normal, invasive ductal carcinoma and ductal carcinoma *in situ* of the breast tissue. *J. Raman Spectrosc.* **2007**, 38:1345–1351.

21. Barman, I., Dingari, N.C., Saha, A., McGee, S., Galindo, L.H., Liu, W., Plecha, D., Klein, N., Dasari, R.R. and Fitzmaurice, M. Application of Raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy. *Cancer Res.* **2013**, 73:3206–3215.

22. Haka, A.S., Shafer-Peltier, K.E., Fitzmaurice, M., Crowe, J., Dasari, R.R. and Feld, M.S. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. *Cancer Res.* **2002**, 62:5375–5380.

23. Haka, A.S., Shafer-Peltier, K.E., Fitzmaurice, M., Crowe, J., Dasari, R.R. and Feld, M.S. Diagnosing breast cancer by using Raman spectroscopy. *Proc. Natl. Acad. Sci. USA.* **2005**, 102:12371–12376.

24. Stone, N. and Matousek, P. Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis. *Cancer Res.* **2008**, 68:4424–4430.

25. Beaumont, T. and Leadbeater, M. Treatment and care of patients with metastatic breast cancer. *Nurs. Stand.* **2011**, 25:49–56.

26. Disibio, G. and French, S.W. Metastatic patterns of cancers: results from a large autopsy study. *Arch. Pathol. Lab. Med.* **2008**, 132:931–939.

27. Brinkley, B.R., Beall, P.T., Wible, L.J., Mace, M.L., Turner, D.S. and Cailleau, R.M. Variations in cell form and cytoskeleton in human breast carcinoma cells *in vitro*. *Cancer Res.* **1980**, 40:3118–3129.

28. Cailleau, R., Olive, M. and Cruciger, Q.V. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro*. **1978**, 14:911–915.

29. Winnard, P.T., Jr., Kluth, J.B. and Raman, V. Noninvasive optical tracking of red fluorescent protein-expressing cancer cells in a model of metastatic breast cancer. *Neoplasia.* **2006**, 8:796–806.

30. Zygourakis, K., Bizios, R. and Markenscoff, P. Proliferation of anchorage-dependent contact-inhibited cells: I. Development of theoretical models based on cellular automata. *Biotechnol. Bioeng.* **1991**, 38:459–470.

31. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, 24:498–520.

32. Jolliffe, I.T. Principal component analysis. (New York: Springer-Verlag). **1986**.

33. Chen, M., Zang, M., Wang, X. and Xiao, G. A powerful Bayesian meta-analysis method to integrate multiple gene set enrichment studies. *Bioinformatics.* **2013**, 29:862–869.

34. Reddy, R.K. and Bhargava, R. Chemometric Methods for Biomedical Raman Spectroscopy and Imaging. In: Matousek P and Morris MD, eds. Emerging Raman Applications and Techniques in Biomedical and Pharmaceutical Fields. (New York: Springer), **2010**, pp:179–213.

35. Kong, K., Rowlands, C.J., Varma, S., Perkins, W., Leach, I.H., Koloydenko, A.A., Williams, H.C. and Notingher, I. Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy. *Proc. Natl. Acad. Sci. USA.* **2013**, 110:15189–15194.

36. Chambers, A.F. MDA-MB-435 and M14 cell lines: identical but not M14 melanoma? *Cancer Res.* **2009**, 69:5292–5293.

37. Hollestelle, A., Nagel, J.H., Smid, M., Lam, S., Elstrodt, F., Wasielewski, M., Ng, S.S., French, P.J., Peeters, J.K., Rozendaal, M.J., Riaz, M., Koopman, D.G., Ten Hagen, T.L., et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res. Treat.* **2010**, 121:53–64.

38. Montel, V., Suzuki, M., Galloy, C., Mose, E.S. and Tarin ,D. Expression of melanocyte-related genes in human breast cancer and its implications. *Differentiation.* **2009**, 78:283–291.

39. Nerlich, A.G. and Bachmeier, B.E. Density-dependent lineage instability of MDA-MB-435 breast cancer cells. *Oncol. Lett.* **2013**, 5:1370–1374.

40. Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell.* **2006**, 10:515–527.

41. Wang, H., Huang, S., Shou, J., Su, E.W., Onyia, J.E., Liao, B. and Li, S.. Comparative analysis and integrative classification of NCI60 cell lines and primary tumors using gene expression profiling data. *BMC Genomics.* **2006**, 7:166.

42. Zhang, Q., Fan, H., Shen, J., Hoffman, R.M. and Xing, H.R. Human breast cancer cell lines co-express neuronal, epithelial, and melanocytic differentiation markers *in vitro* and *in vivo*. *PLoS One.* **2010**, 5:e9712.

43. Huang, Z., The, S.K., Zheng, W., Lin, K., Ho, K.Y., The, M. and Yeoh, K.G. *In vivo* detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy. *Biosens. Bioelectron.* **2010**, 26:383–389.

44. Lieber, C.A. and Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl. Spectrosc.* **2003**, 57:1363–1367.

45. Gabius, H.J., Ciesiolka, T., Kunze, E. and Vehmeyer, K. Detection of metastasis-associated differences for receptors of glycoconjugates (lectins) in histomorphologically unchanged xenotransplants from primary and metastatic lesions of human colon adenocarcinomas. *Clin. Exp. Metastasis.* **1989**, 7:571–584.

46. Price, J.E., Tarin, D. and Fidler, I.J. Influence of organ microenvironment on pigmentation of a metastatic murine melanoma. *Cancer Res.* **1988**, 48:2258–2264.

47. Hass, R. and Bertram, C. Characterization of human breast cancer epithelial cells (HBCEC) derived from long term cultured biopsies. *J. Exp. Clin. Cancer Res.* **2009**, 28:127.

48. Shay, J.W. and Wright, W.E. Tissue culture as a hostile environment: identifying conditions for breast cancer progression studies. *Cancer Cell.* **2007**, 12:100–101.

49. Visconti, R., Della Monica, R. and Grieco, D. Cell cycle checkpoint in cancer: a therapeutically targetable double-edged sword. *J. Exp. Clin. Cancer Res.* **2016**, 35:153.

50. Rota, L.M., Lazzarino, D.A., Ziegler, A.N., LeRoith, D. and Wood, T.L. Determining mammosphere-forming potential: application of the limiting dilution analysis. *J. Mammary Gland Biol. Neoplasia.* **2012**, 17:119–123.

51. Abounit, S. and Zurzolo, C. Wiring through tunneling nanotubes- -from electrical signals to organelle transfer. *J. Cell Sci.* **2012**, 125:1089–1098.

52. Antanaviciute, I., Rysevaite, K., Liutkevicius, V., Marandykina, A., Rimkute, L., Sveikatiene, R., Uloza, V., Skeberdis, V.A. Long-distance communication between laryngeal carcinoma cells. *PLoS One.* **2014**, 9:e99196.

53. Lim, Y.S. and Tang, B.L. Intercellular organelle trafficking by membranous nanotube connections: a possible new role in cellular rejuvenation? *Cell Commun. Adhes.* **2012**, 19:39–44.

54. Tyler-McMahon, B.M., Boules, M. and Richelson, E. Neurotensin: peptide for the next millennium. *Regul. Pept.* **2000**, 93:125–136.

55. Trams, E.G., Fales, H.A. and Gal, A.E. S-palmityl pantetheine as an intermediate in the metabolism of palmityl Coenzyme A by rat liver plasma membrane preparations. *Biochem. Biophys Res. Commun.* **1968**, 31:973–976.

56. Nakamura, M., Onodera, T. and Akino, T. Characteristics of phospholipids in human lung carcinoma. *Lipids.* **1980**, 15:616–623.

57. Umemori, Y., Ohe, Y., Kuribayashi, K., Tsuji, N., Nishidate, T., Kameshima, H., Hirata, K. and Watanabe, N. Evaluating the utility of N1, N12-diacetylspermine and N1, N8-diacetylspermidine in urine as tumor markers for breast and colorectal cancers. *Clin. Chim. Acta.* **2010**, 411:1894–1899.

58. Abramczyk, H., Surmacki, J., Kopec, M., Olejnik, A.K., Kaufman-Szymczyk, A. and Fabianowska-Majewska, K. Epigenetic changes in cancer by Raman imaging, fluorescence imaging, AFM and scanning near-field optical microscopy (SNOM). Acetylation in normal and human cancer breast cells MCF10A, MCF7 and MDA-MB-231. *Analyst.* **2016**, 141:5646–5658.

59. Abramczyk, H., Surmacki, J., Kopec, M., Olejnik, A.K., Lubecka-Pietruszewska, K. and Fabianowska-Majewska, K. The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue. *Analyst.* **2015**, 140:2224–2235.

60. Manciu, F.S., Ciubuc, J.D., Parra, K., Manciu, M., Bennet, K.E., Valenzuela, P., Sundin, E.M., Durrer, W.G., Reza, L. and Francia, G. Label-Free Raman Imaging to Monitor Breast Tumor Signatures. *Technol. Cancer Res. Treat.* **2016**, 16:461-469.

61. Nieva, C., Marro, M., Santana-Codina, N., Rao, S., Petrov, D. and Sierra, A. The lipid phenotype of breast cancer cells characterized by Raman microspectroscopy: towards a stratification of malignancy. *PLoS One.* **2012**, 7:e46456.

62. Surmacki, J., Brozek-Pluska, B., Kordek, R. and Abramczyk, H. The lipid-reactive oxygen species phenotype of breast cancer. Raman spectroscopy and mapping, PCA and PLSDA for invasive ductal carcinoma and invasive lobular carcinoma. Molecular tumorigenic mechanisms beyond Warburg effect. *Analyst.* **2015**, 140:2121–2133.

63. Carreau, A., El Hafny-Rahbi, B., Matejuk, A., Grillon, C. and Kieda, C. Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. *J. Cell Mol. Med.* **2011**, 15:1239–1253.

64. Sun, X., Voloboueva, L.A., Stary, C.M. and Giffard, R.G. Physiologically normal 5% O$_2$ supports neuronal differentiation and resistance to inflammatory injury in neural stem cell cultures. *J. Neurosci. Res.* **2015**, 93:1703–1712.

Chapter 6. Summary and future work

6.1 Summary

This thesis reports multiple applications of spontaneous Raman spectroscopy from protein aggregation analysis to detection of early-stage metastatic cancer in the bone. We have also employed SERS by harnessing specialized nano-structured substrates to achieve ultrasensitive bioanalyte detection. In our investigations, machine learning and multivariate analysis has been extensively utilized to gain mechanistic insights into protein aggregation and cancer metastasis. Besides, fluorescence microscopy, UV-Vis absorption spectroscopy and finite-difference time-domain simulation are also included in the dissertation to support our findings from Raman spectroscopic studies.

In our first spectroscopic study, we developed a support vector machine-based decision model that offers fast, accurate prediction for a wide range of protein aggregation, particularly at the low aggregation levels, based on the subtle, but consistent, differences in their spectra, which are otherwise invisible to gross visual inspection. Furthermore, in blinded experiments, the model shows the ability to precisely differentiate between aggregation levels in mAb samples pre- and post-isothermal incubation, where the latter is representative of protein degradation achieved under long-term storage conditions. The high degree of accuracy achieved in prediction in these stressed/unstressed samples that exhibit only small differences in aggregation levels is particularly encouraging. This newly developed method takes less than three minutes to finish each measurement – much faster than standard HP-SEC measurements. This study presents the power of Raman spectroscopy to rapidly recognize protein particle formation during dynamic

biomanufacturing. In addition to offering a new tool in support of development and manufacturing of protein drugs and biosimilars, this method may also benefit regulatory authorities by helping develop improved guidance parameters for manufacturing of safe and effective protein therapeutics.

Although spontaneous Raman spectroscopy could offer wealth of molecular information of the sample, the low probability of Raman scattering events hinders the acquisition of high signal-to-noise ratio signals. Hence, we use SERS to enhance the Raman signal. In the second study, we show a novel dual-modality nanostructure design that combines LSPR and SERS to detect the variation of RI and obtain unique molecular fingerprints. By profiling two distinct resonance peaks, we achieve sensitive and robust detection of RI charges due to molecular concentration variations. Also, clear SERS spectral signal is obtained from solutions with low analyte concentration, where no spontaneous Raman signal could be observed. Through further optimization of the structure design and corresponding fabrication procedures, we envision this dual-modality nanostructure could be leveraged for translating molecular markers into serum assays for accurate disease screening.

By virtue of its exquisite biochemical specificity, Raman spectroscopy also allows the visualization of complex molecular heterogeneity directly from cells and tissues and has been extensively used in biomedical studies to quantify the unique vibrational modes of molecules within its native context. In our third endeavor, we demonstrate that Raman spectroscopy is able to detect biomolecular changes in bones associated with early cancer metastasis - much earlier than X-ray imaging and without *a priori* imaging or pathological knowledge of lesion location. This research paves the way for continuous monitoring of

such sites to circumvent pathological fractures. This approach could transform the exciting and emerging area of research that strives to elucidate the interplay between biochemical signals and morphologic features in the bone microenvironment, and precisely pinpoint alterations as a function of metastasis onset and progression.

Finally, in chapter 5, we uncover a few fundamentally new distinctions between isogenic metastatic cancer cells and primary tumor cells by analyzing their vibrational spectra. Our results, in particular, suggest that alterations in lipogenesis processes along with the types of lipids made and utilized could be part of the determinant adaptations that define the specific metastatic growth in an organ. The importance of the collective findings is that we can extend the use of Raman micro-spectroscopy as a non-invasive method to differentially characterize organ-specific metastatic lesions and thus, aid in pathological assessments of efficacious personalized therapies for organ-site-specific treatment of metastatic breast cancer. Equally importantly, our work reveals the important morphological, growth distinctions and complementary overlap between the metabolite data set and Raman spectroscopic signatures which support our hypothesis that each metastatic site's unique environment confers singular molecular attributes to the tumor cells despite their isogenic profile.

6.2 Future work

While several important advances have been made during the course of this dissertation, further investigations are needed to translate these research findings to the clinic as well as an in-line analytical tool for biopharmaceutical development.

For the latter, we observed that most mAb samples exhibit different aggregation mechanisms and, hence, the particle formation process renders different spectral signatures. In the presented study, we focused on a specific kind of ADC sample due to product-specific processes on individual manufacturing pipelines in the pharmaceutical industry. The established regression model is also specifically optimized for the investigated samples. Hence, to further establish the generalizability of the method, more types of mAb products should be included in future spectral measurements while sample-specific protein aggregation prediction models could be set up for individual production pipelines.

Additionally, our study of breast cancer-colonized bone alterations clarifies the power of spontaneous Raman spectroscopy to detect early, subtle biochemical changes. In order to make our results benefit patients, it is necessary to link the bone strength with tumor progression and Raman spectra quantitively. Therefore, a series of mechanical measurements will need to be performed and corresponding prediction models trained to figure out the intrinsic relationship between these stages of tumor progression.

Finally, the diverse Raman spectroscopic signatures and metabolic features in different organ-specific IMBC cell lines provide solid rationale for establishment of smart, site-specific therapeutics. This is an important consideration as routine clinical treatment decision options for distant metastatic breast cancer have historically relied in part on an evaluation of a select few biomarkers found during assessment of the primary tumor. In the future, our observations on this unique set of organ-specific metastatic breast cancer cell lines will be directly validated using patient samples for clinical relevance. The data that will be generated will be immensely useful to the tailor ongoing clinical trials, whereby targeted smart therapeutic strategy can be employed. We may also be able identify new

therapeutic options to treat specific breast cancer metastatic lesions using a defined set of FDA approved oncological drugs. The rationale of which FDA drugs to select for testing will then be firmly based on those drugs mechanistically known to be effective against the pathways uncovered.

Curriculum Vita

Date of Birth: October 28, 1990

Place of Birth: Beijing, China

Education

- Ph.D., Johns Hopkins University, Baltimore, Maryland   09/2015-08/2019

- M.Eng., Johns Hopkins University, Baltimore, Maryland   09/2013-05/2015

- B.S., Beihang University, Beijing, China   09/2009-07/2013

Publications

1. **Zhang, C.**, Springall, J., Wang, X, Barman, I. Rapid, quantitative determination of aggregation and particle formation for antibody drug conjugate therapeutics with label-free Raman spectroscopy. *Analytica Chimica Acta* 2019, 1081, 138-145.

2. **Zhang, C.**,[#] Paria, D.,[#] Semancik, S., Barman, I. Composite-scattering plasmonic nanoprobes for label-free, quantitative biomolecular sensing. *Small* 2019, 1901165.

3. **Zhang, C.**, Siddhanta, S., Li, Y, Zheng, C., Barman, I. Label-free nanoscopic plasmonic sensing of membrane specific prostate cancer biomarkers. (In preparation)

4. Junjuri, R., **Zhang, C.**, Barman, I., Kumar, G.M. Identification of post-consumer plastics using laser-induced breakdown spectroscopy. *Polymer Testing* 2019, 76, 101-108.

5. Karandikar, S., **Zhang, C.**, Meiyappan, A., Barman, I., Finck, C., Srivastava, P.K. and Pandey, R., Reagent-free and rapid assessment of T cell activation state using

diffraction phase microscopy and deep learning. *Analytical Chemistry* 2019, 91(5), 3405-3411.

6. Sperati, C.J., **Zhang, C.**, Delsante, M., Gupta, R., Bagnasco, S. and Barman, I. Raman Spectroscopy for the Diagnosis of Intratubular Triamterene Crystallization. *Kidney International Reports* 2018, 3(4), 997-1003.

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

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

9. Pandey, R., Singh, S.P., **Zhang, C.**, Horowitz, G.L., Lue, N., Galindo, L., Dasari, R.R. and Barman, I. Label-free spectrochemical probe for determination of hemoglobin glycation in clinical blood samples. *Journal of biophotonics* 2018, e201700397.

10. Winnard Jr, P.T., **Zhang, C.**, Vesuna, F., Kang, J.W., Garry, J., Dasari, R.R., Barman, I. and Raman, V. Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes. *Oncotarget* 2017, 8(12), 20266–20287.

11. Pandey, R., Paidi, S.K., Valdez, T.A., **Zhang, C.**, Spegazzini, N., Dasari, R.R. and Barman, I. Noninvasive Monitoring of Blood Glucose with Raman Spectroscopy. *Acc. Chem. Res*. 2017, 50(2), 264-272.

12. Huang, Z., Siddhanta, S., **Zhang, C.**, Kickler, T., Zheng, G. and Barman, I. Painting and heating: A nonconventional, scalable route to sensitive biomolecular analysis with plasmon-enhanced spectroscopy. *Journal of Raman Spectroscopy* 2017, *48*(10), 1365-1374.

13. Haka, A.S., Sue, E., **Zhang, C.**, Bhardwaj, P., Sterling, J., Carpenter, C., Leonard, M., Manzoor, M., Walker, J., Aleman, J.O. and Gareau, D. Noninvasive detection of inflammatory changes in white adipose tissue by label-free Raman spectroscopy. *Analytical chemistry* 2016, 88(4), 2140-2148.

14. Myakalwar, A.K., Spegazzini, N., **Zhang, C.**, Anubham, S.K., Dasari, R.R., Barman, I. and Gundawar, M.K. Less is more: Avoiding the LIBS dimensionality curse through judicious feature selection for explosive detection. *Scientific reports* 2015, *5*, 13169.

Presentation & Posters

1. **Zhang, C.**, Barman, I. Raman spectroscopy of isogenic breast cancer cells derived from organ-specific metastases reveals distinct biochemical signatures. *SLAS.* January, 2018, San Diego.

2. **Zhang, C.**, Barman, I. Early determination of molecular alterations in breast cancer-colonized bone with Raman spectroscopy. *SPIE BiOS.* January, 2018, San Francisco.

3. **Zhang, C.**, Siddhanta, S., Zheng, C., Barman, I. Probing nanoscopic cell surface areas for rapid and label-free plasmon enhanced Raman detection. *Biomedical Optics.* April, 2016, Fort Lauderdale.

4. **Zhang, C.**, Winnard Jr., P.T., Raman, V., Barman, I. Label-free identification of unique metastatic organ-specific human breast cancer signatures featuring Raman

spectroscopy. *2nd International Conference on Label-Free Technologies.* March, 2015, Boston.