RAMAN SPECTROSCOPIC ANALYSIS OF FATTY ACIDS IN TUMOR MICRO- AND MACROENVIRONMENTS IN BREAST CANCER

BY

SIXIAN YOU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Bioengineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Adviser:

Professor Stephen A. Boppart

Abstract

Fatty acids play essential roles in the growth and metastasis of cancer cells. To facilitate their avid growth and proliferation, cancer cells not only alter the fatty acid synthesis and metabolism intracellularly and extracellularly, but also in the macroenvironment via direct and indirect pathways. This thesis reports that by using Raman micro-spectroscopy, an increase in the production of polyunsaturated fatty acids (PUFAs) was identified in both cancerous and normal appearing breast tissue obtained from breast cancer patients and tumor-bearing rats. By minimizing confounding effects from mixed chemicals and optimizing the signal-to-noise ratio of Raman spectra, a large-scale transition from monounsaturated fatty acids to PUFAs was observed in the tumor while only a small subset of fatty acids transitioned to PUFAs in the tumor micro- and macroenvironment. These findings demonstrate the power of this spectroscopic analysis, and may provide new insights into the macroenvironmental regulations in breast cancer.

Acknowledgments

I would like to thank my advisor Professor Stephen Boppart for his continuous support, encouragement, and valuable insights throughout the learning process of this thesis work. It has been a truly enjoyable experience to work in the Biophotonics Imaging Laboratory (BIL).

I also would like to thank Dr. Marina Marjanovic, Dr. Youbo Zhao, Dr. Yuan Liu, and Dr. Haohua Tu for their patience and generosity to share their immense expertise and valuable life lessons with me. I would also like to express my gratitude to Eric Chaney for helping me acquire all the biological samples used in this work, as well as Darold Spillman for his administrative support and technical assistance. Special thanks to all of my friends and colleagues in BIL for being supportive and caring, and making this lab the best place to work.

Finally, I would like to thank my family as well as my best friend, Lin, for always being supportive and understanding wherever I go and whatever I do.

Table of Contents

1   Introduction

1.1   Tumor micro- and macroenvironment in breast cancer

Despite the rapid technological advance in cancer diagnosis and treatment, breast cancer continues to persist as the second most common cancer, as well as the second most common cause of cancer-related deaths among women in the United States (1). There remains a need for a better understanding of breast cancer development and for more effective diagnostic and therapeutic strategies that can help to reduce the burden of breast cancer.

Among the best-studied areas of breast cancer is the alterations that occur in cancer cells. Despite their genotypic diversity, it has been shown that at least three pronounced phenotypes exist in all aggressive cancers. One of them is related to the Warburg effect, which describes the high-rate consumption of glucose and production of lactic acid within cancer cells even under aerobic conditions (2,3). A second hallmark is increased protein synthesis and more active DNA synthesis driven by vigorous and constant proliferation and dissemination (4). A third hallmark is an increased *de novo* fatty acid synthesis, which is critically related to energy metabolism, cell membrane formation, and signaling events (5).

Less well studied, although supported by a wide range of imaging and genetic studies, is the active interaction between the tumor microenvironment and the tumor (6,7). Stromal cells, including fibroblasts (8), macrophages (9), adipocytes (10), and vascular cells (11), have been shown to be recruited to the tumor sites to generate a tumorigenic microenvironment. It also has been shown that non-cellular components in the tumor microenvironment, such as the extracellular matrix (ECM), have direct and indirect impacts on cancer and stromal cells, thereby facilitating migration and invasion (12–14). It is now widely acknowledged that the interactions between

tumor cells and their local environment play important roles in cancer development (3,6,15,16). These studies have provided invaluable insights into the development, diagnosis, and treatment of cancer (17–19).

More recently, however, scientists and clinicians have increasingly begun to expand their investigations from the local tumor microenvironment to a more global tissue macroenvironment (tissue at a larger distance from the primary tumor), appreciating the active participation of a complex array of host factors in tumor progression (20,21). As cancer cells are vigorously and constantly replicating and disseminating themselves, they impose a larger requirement to maintain their viability compared to non-proliferating (differentiated) cells (22). These avidly proliferating cells have to create a tumor-supportive environment locally and systematically to meet the demanding requirements of resources and also protect themselves from possible immune attacks. It has been shown that the macroenvironmental regulation of cancer occurs at the genetic, proteomic, and metabolic levels to elicit carcinogenesis, metastasis, and progression (20,23,24). For instance, previous reports identified genetic alterations not only in the stromal cells close to breast carcinomas, but also in the tumor-free and normal appearing skin of the patients with breast cancer (20,25). Understanding how tumors interact with the macroenvironment is likely to have a major impact on our understanding of the complex mechanisms underlying cancer development, as well as establish links between cancer and metabolic diseases such as obesity and diabetes. Moreover, these new insights could also contribute to new therapeutic targets for cancer in general, and breast cancer in particular.

Despite such importance, the quantitative evaluation of changes in tissues remote from tumor sites remains largely unexplored. There continues to be a need for quantitative characterization of the tumor macroenvironment for potential use in biological and clinical studies

that are devoted to treating cancer-related metabolic disease and developing macroenvironment-targeted therapeutic treatments.

1.2   Lipid metabolism in breast cancer

As noted, there is growing evidence that the metabolism of cancer cells is distinct from that of cells in normal tissue due to their different metabolic requirements (22,26). Survival of a mammalian organism highly depends on the precise control of individual cell proliferation and strict regulation of nutrient up-take. In normal differentiated mammalian cells, proliferation is only possible when exposed to specific growth factors and thus aberrant individual cell proliferation is prevented. In contrast, such growth factor dependence does not exist for cancer cells. The ultimate goal of cancer cells is to survive, grow, and proliferate irrespective of whether they are endangering the survival of their host. In order to meet the usual demands of such rapid growth and proliferation of these cancer cells, several significant signaling pathways are altered to activate the uptake and metabolism of nutrients that fuel cell growth and promote cell dissemination.

Among all the metabolic pathways, lipid metabolism has recently gained much prominence for its diverse roles in cell membrane formation, energy storage, and signaling pathways (27). The term lipid includes a diverse group of hydrophobic compounds that are mostly composed of fatty acids bound to a polyalcohol with an ester bond. Some of the lipids are major constituents of all biological membranes, such as phospholycerides, sterols, and sphingolipids. Triacylglycerol is another form of lipids and an important substrate for energy metabolism. In addition, some of the lipids, such as cholesterol, serve as signaling molecules in the form of messengers and hormones to sustain homeostasis.

Given the important physiological roles of lipids at both cellular and organismal levels, it is not surprising to see that alterations in lipid metabolism profoundly affect the regulation of

3

proliferation, differentiation, apoptosis, autophagy, inflammation, motility, and membrane homeostasis (28). One of the first and most important breakthroughs of studies on lipid metabolism in cancer is the discovery of *de novo* lipogenesis in cancer cells (29). Most normal cells tend to use exogenously derived (dietary) fatty acids in a well-nourished environment. However, irrespective of the levels of circulating lipids, cancer cells tend to resort to endogenously synthesized fatty acids, which normally is only very active during embryogenesis and in fetal lungs of healthy individuals. Such upregulation of fatty acid synthase is not only a secondary consequence of carcinogenesis, but more importantly, it is directly selected for the growth and survival advantage it provides via multiple mechanisms (30).

Increased cholesterol biosynthesis is another important event in dysregulated lipid metabolism induced by breast cancer. Cholesterol is an important constituent of biological membranes due to its ability to modulate the fluidity of the lipid bilayer and facilitate the signaling events at the membrane (27). In most cells, free cholesterol is maintained at a relatively constant level by several homeostatic processes (31). In contrast, such homeostasis is disrupted in avidly proliferating breast cancer cells. It has been difficult, however, to understand how this is enabled by cancer cells, given the complexity and redundancy of the homeostatic processes. Nevertheless, there are plenty of data suggesting that cholesterol and/or its derivative acts as a mediator of the effects of dyslipidemia and/or obesity in breast cancer pathogenesis (31).

Another important aspect of studies on altered lipid metabolism is the analysis of fatty acid composition, which is a simple and direct reflection of the main alterations in the lipid synthesis and metabolism. It has been shown that fatty acid composition in normal adipose tissue is a good indicator of long-term intake of different fatty acids as normal cells rely heavily on circulating dietary lipids (32). However, in cancer cells, 90% of triglycerides were synthesized *de novo*. As a

result, fatty acid composition in cancer cells is controlled and regulated by their special metabolic requirements rather than just the dietary input. Therefore, the study of fatty acid compositional changes in cancer subjects is far more than a study of dietary fatty acid input as in the case of normal subjects, but rather a study of the influence from cancer cells on lipid metabolism. For instance, a three- to four-fold increase in the content of phospholipids has been reported in the breast cancer tissue compared with the surrounding normal breast adipose tissue of the same women (33). This significant increase in the production of phospholipids is independent from dietary input difference and very likely to be a result from high demands of biological membrane formation by rapidly proliferating cancer cells.

Much recent work on lipid metabolism and breast cancer has been focused on the understanding of the complex relationship between obesity and cancer. As a common form of metabolic dysfunction, obesity is positively associated with higher risk for developing breast cancer, poorer prognoses, and diminished responses to chemotherapy. The most recent data show that obesity has overtaken smoking as the number one preventable cause of cancer-related deaths in the United States. This increased cancer risk could be induced by the systemic metabolic dysregulations that accompany obesity, including elevated levels of cholesterol and other lipids, and high levels of circulating insulin and hormones. However, it has remained largely elusive regarding which host factors are the key factors mediating the obesity-cancer link, which is a major challenge in understanding the complex mechanisms linking obesity and cancer.

1.3 Polyunsaturated fatty acids in breast cancer

Fatty acids can be classified as saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids (PUFAs) based on the number of unsaturated bonds in the fatty acid chain, as shown in Figure 1.1. The unsaturated bonds in the long chain improve the flexibility and

permeability of biological membranes, allowing the transportation of nutrients and waste products. Although PUFAs only account for approximately 12% of the fatty acid content in normal and cancerous human adipose tissue (34), PUFAs have recently gained more attention due to their important role in modulating membrane properties, receptor signaling, and the expression of several key genes in adipogenesis (35).

PUFAs can be categorized into two groups: n-3 fatty acids and n-6 fatty acids, depending on the location of the first double bond relative to the omega end of the fatty acid chain, as shown in Figure 1.1. For instance, for an n-3 fatty acid, the first double bond occurs between the third and fourth carbon atom away from the omega end. Because linoleic (n-6) and α-linoleic (n-3) acids cannot be synthesized *de novo* by mammals, both types of fatty acids are defined as essential fatty acids and have to be obtained through diet. The n-3 fatty acids can be obtained from high-fat fish such as salmon and tuna, whereas n-6 fatty acids are mainly obtained through organ meats and vegetable oils, such as soybean and corn oil (36). Both of these are stored in esterified form in phospholipids of biological membranes for cells and organelles, or in phospholipids and glycerides found in lipid droplets (37,38). Once released from an esterified form, these fatty acids are further metabolized to eicosanoids and other autacoids. Because the same series of enzymes are used for the metabolism of n-3 and n-6 fatty acids, an excess of one fatty acid will cause a significant decrease in the metabolism of the other due to the competition of the same enzymes (35).
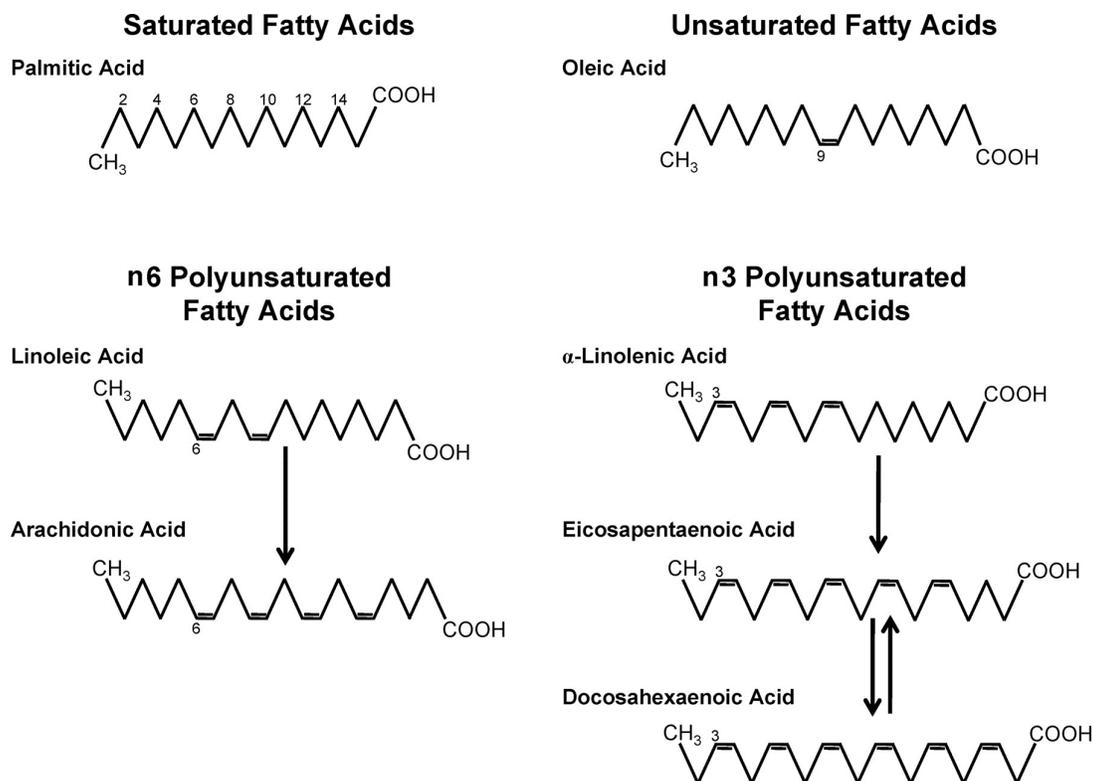
Figure 1.1 Chemical structure of saturated, monounsaturated, and n-3 and n-6 polyunsaturated fatty acids (39)**.**

Although the mechanisms of PUFAs affecting biological systems are not completely understood, it is well acknowledged that PUFAs plays several key roles in cellular functions. First of all, PUFA content of biological membranes significantly alters membrane fluidity and permeability, thereby influencing vesicle budding and fusion, and protein function and trafficking (35). With an increase in the PUFA levels, membrane fluidity increases greatly as conformational states can be rapidly changed with substantially more flexible PUFA acyl chains. Membrane fluidity also increases with the n-6 fatty acid composition since n-6 arachidonic (AA; 20:4) acyl chains have been shown to be more deformable and form softer bilayer than n-3 docosahexaenoic (DHA; 22:6) acyl chains (40). Secondly, PUFA-derived messengers showed anti-inflammatory (n-3) or pro-inflammatory (n-6) effects. The n-6 derivatives promote inflammation by increasing vascular permeability and producing inflammatory cytokine whereas

7

the n-3 derivatives reduce inflammation by suppressing the synthesis of inflammatory cytokines and chemokines, and elevating the production of anti-inflammatory autacoids (41,42). Finally, PUFAs affect cellular metabolism, growth, and differentiation through modulating gene expression, especially genes involved in lipid metabolism. Genes related to more active fatty acid transport and metabolism tend to be upregulated by PUFAs, such as acyl CoA synthetase (43) and fatty acid transport protein (44), whereas lipogenic genes such as fatty acid synthase (45) and stearoyl CoA desaturase 1 (46) tend to be suppressed by PUFAs. However, no certain conclusion has been drawn on the functions of PUFAs because fatty acid metabolism is found to be site-specific and such gene regulations vary among different adipose tissues.

Given the important roles of PUFAs in cellular metabolism, growth, and differentiation, extensive studies have been trying to clarify the cause and effect relationships between PUFAs and cancer progression. However, these studies have produced equivocal results, with many finding no correlation, some implicating a protective effect, and others indicating PUFAs as a significant risk factor. Some of these discrepancies can be attributed to the differences between n-3 and n-6 fatty acids or the difference of the impact of PUFAs on different types of cancer. Indeed, n-6 PUFAs have been shown to promote tumorigenesis via increasing circulating estrogenic compound levels (47), cell proliferation rate (48), and inflammation (49), while n-3 PUFAs have been shown to be responsible for suppressing tumor growth by decreasing circulating estrogenic compound levels (47), breast cancer cell growth (50), and HER2 expression (51). One recent cohort study demonstrated that patients with breast cancer had significantly lower n-3 fatty acids but seemingly higher average n-6 fatty acids. It was also demonstrated, in a cancer cell study, that PUFA content is positively associated with higher metastatic ability in breast cancer cells (52). These observations strongly suggest a link between PUFA levels and breast cancer development.

8

1.4 Summary

Altered lipid metabolism is a hallmark of breast cancer (26,27,30). An increase in PUFA levels has been positively associated with the aggressiveness of cancer cell lines (52) and the promotion of tumorigenesis (53). It has been shown that higher PUFA levels increase the risk of metastasis in cancer patients by increasing estrogen levels (47,53), by the expression of cancer-promoting genes such as PAI-1 (54,55), and by the adherence of circulating tumor cells to blood vessel walls and remote organs (56). Although it is known that lipid metabolism is drastically altered in cancer cells, it is not known whether such effects occur in more distant tissue sites, i. e. the macroenvironment. Evidence from other lipid-related studies strongly suggests that metabolic diseases such as obesity and diabetes can interfere with several lipogenic regulatory pathways, which in turn may give rise to an increased predisposition for cancer (57,58). Despite the importance of these studies to the development of new interventions and therapeutics, further investigations are hindered by the lack of reliable biochemical signatures for the tumor macroenvironment. Hence, the goal of this thesis is to discover biochemical signatures that reveal unusual activity in lipid metabolism in the tumor micro- and macroenvironment in breast cancer.

2 Raman spectroscopy

2.1 Raman spectroscopy: theory

Raman spectroscopy is a noninvasive, label-free, and chemical-specific technique that probes molecular structure based on the scattering interaction between incident photons and targeted matter. When light is scattered from a molecule, a large proportion of the photons are elastically scattered and end up with the same energy (i.e. frequency) as the incident photons, whereas a very small fraction of the photons (approximately 1 out of ten million photons) are scattered at a lower or higher frequency, which is called inelastic scattering (59,60). Raman scattering is an inelastic scattering process in which the incident photon interacts with the electronic cloud of the molecule and results in an energy transfer from or to various molecular vibrational modes, accompanied by a frequency shift of the incident photon (61).

If a molecule is in the vibrational ground state before the irradiation of incident photons, its final state will be either the same vibrational ground state, called Rayleigh scattering or elastic scattering, or an excited state, called Stokes Raman scattering. In Stokes Raman scattering, abiding by the law of energy conservation, a molecule absorbs an incident photon of frequency v_i and emits a Raman photon of lower frequency at v_R simultaneously, accompanied by its transition from the ground state to an excited state. In this case, the energy difference between the vibrational ground state and the final excited state is calculated by:

$$\Delta E = h(v_i - v_R) \tag{2.1}$$

where h is Planck's constant (units $J \cdot s$). The vibrational energy gained by the molecule is ultimately dissipated as heat. On the other hand, if the molecule begins in an excited vibrational mode, it will transfer energy to the photon and return to the vibrational ground state, called anti-

Stokes Raman scattering. In anti-Stokes Raman scattering, the incident photon of frequency v_i gains energy from the molecule and results in a scattered photon with higher frequency v_R. The energy difference in this case is calculated by:

$$\Delta E = h(v_R - v_i).$$ (2.2)

As shown by Eqn. (2.1) and Eqn. (2.2), anti-Stokes and Stokes Raman scattering carry identical frequency information about the molecule. According to a Boltzmann distribution, the ratio of the populations of the excited state to that of the ground state should be very low at room temperature. Therefore, Stokes Raman scattering is usually the dominant transition measured in Raman spectroscopy.

According to Eqn. (2.1) and Eqn. (2.2), the vibrational modes of a certain molecular structure can be assessed by calculating the difference between the frequency of the incident photon and the Raman-scattered photon. Thus, by recording the frequency shifts of the incident photon, the structure of the probed molecule can be revealed. In Raman spectroscopy, this idea is implemented by using a monochromatic laser as an excitation source and measuring the spectrum of the scattered light. In order to understand the physical meaning of Raman spectral bands, the classical electromagnetic field description of Raman spectroscopy can be used:

$$P = \alpha E$$ (2.3)

where P is the dipole moment induced in the molecule, α is the polarizability of the molecule, and E is the applied electric field. When the polarizability of a molecule stays unchanged, the induced dipole emits light at the same frequency as the incident photon. Considering the fluctuation with time of the incident electromagnetic wave, the induced electric dipole moment can be expressed as:

$$P = \alpha E_0 \cos 2\pi v_i t \tag{2.4}$$

where E_0 is the intensity and v_i is the frequency of the applied electric field. However, in Raman scattering, the polarizability of a molecule is changed by its molecular vibrations. Using small perturbation theory, the polarizability can be approximated as:

$$\alpha = \alpha_0 + (\frac{\partial \alpha}{\partial Q_k})_0 Q_k \tag{2.5}$$

in which α_0 is the original polarizability, and Q_k describes the normal mode of vibration in the molecule at frequency v_k with amplitude $Q_k^{\,0}$:

$$Q_k = Q_k^{\,0} \cos 2\pi v_k t . \tag{2.6}$$

When inserting Eqn. (2.5) and Eqn. (2.6) into Eqn. (2.4), it can be seen that the polarizability takes on the form:

$$P = \alpha_0 E_0 \cos 2\pi v_i t + E_0 Q_k^{\,0} (\frac{\partial \alpha}{\partial Q_k})_0 \cos 2\pi v_i t \cos 2\pi v_k t . \tag{2.7}$$

Applying common trigonometric relations, Eqn. (2.7) can be written as:

$$P = \alpha_0 E_0 \cos 2\pi v_i t + \frac{1}{2} E_0 Q_k^{\,0} (\frac{\partial \alpha}{\partial Q_k})_0 [\cos 2\pi (v_i + v_k)t + \cos 2\pi (v_i - v_{)k} t]. \tag{2.8}$$

The first term is Rayleigh scattering, the second term is the anti-Stokes Raman scattering, and the last term is Stokes Raman scattering. It can be readily seen from Eqn. (2.8) that in order to be Raman-active, the polarizability of the electron cloud must be changed. Thus, the Raman selection rule is given by:

$$\frac{\partial \alpha}{\partial Q_k} \neq 0 . \tag{2.9}$$

Even when the Raman selection rule is satisfied, the Raman signals are still very weak since the Raman signal arises from only an extremely small fraction of the incident photons. The Raman scattered intensity is given by:

$$I \cong v_i^4 I_0 N f(\alpha^2)$$

(2.10)

in which I_0 is the intensity of the incident light, N is the number of scattering molecules, and α is the polarizability of the sample. The term $f(\alpha^2)$ is a function describing the square of the polarizability derivative, i. e. the square of the induced dipole moment, which is proportional to the scattering intensity (62). Eqn. (2.10) is consistent with the Raman selection rule that the Raman band will be close to zero if a vibration does not cause an observable change in the polarizability. This equation also points out that the intensity of Raman scattering depends on the fourth power of the frequency of the incident light. Thus, the choice of the frequency of the excitation source largely determines the signal-to-noise (SNR) of the system.

2.2   Raman spectroscopy: instrumentation

2.2.1   Considerations of Raman-independent signals

It is important to note that the Raman signal is accompanied by strong Raman-independent background signals, including Rayleigh scattering, absorption, and fluorescence, as shown in the Figure 2.1. A typical cross-section for absorption, fluorescence, Rayleigh scattering, and Raman scattering is approximately 10^{-16}, 10^{-18}, 10^{-28}, and 10^{-30} cm^2, respectively. Rayleigh scattering and Raman scattering can vary significantly in different circumstances because both of them depend strongly upon the excitation frequency and the effect of the absorption band. Given the significantly smaller Raman scattering cross-section, these Raman-independent signals have to be taken into consideration in order to retrieve the extremely weak Raman signals.
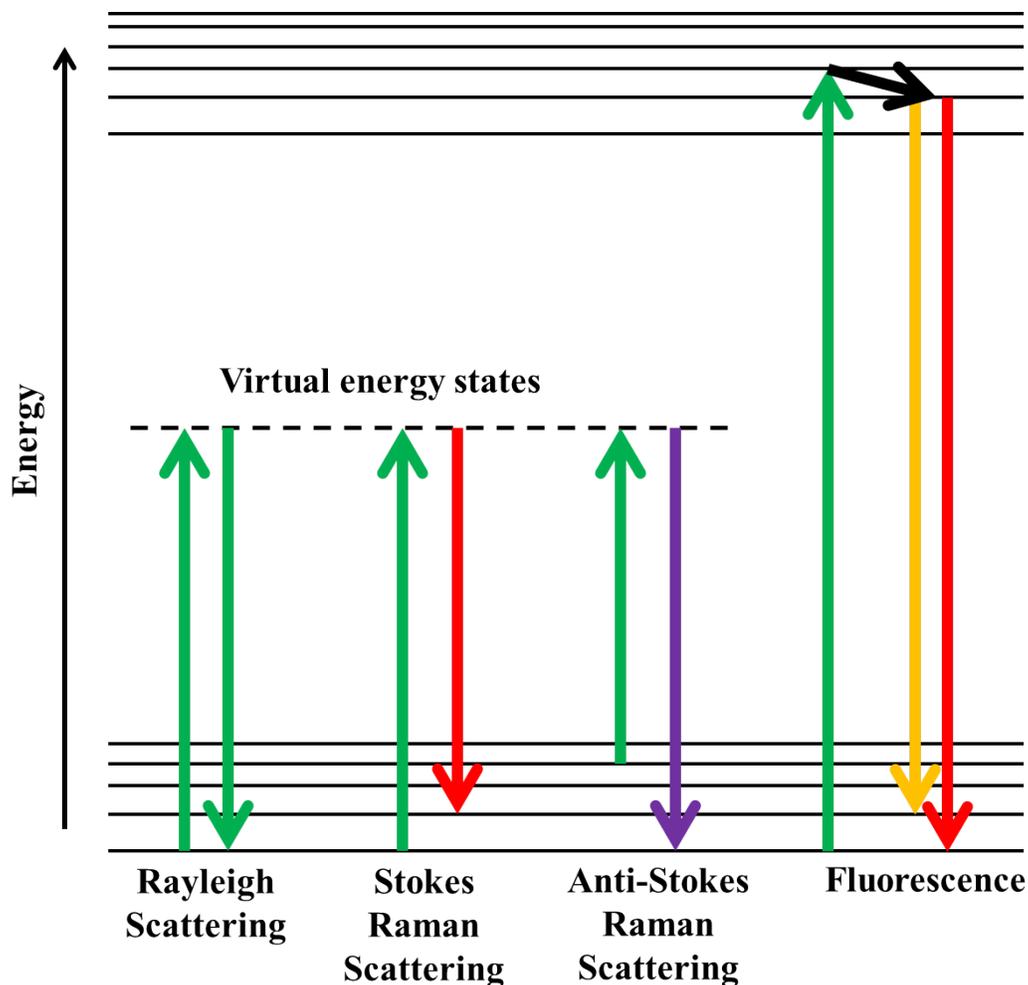
Figure 2.1 Jablonski diagrams depicting the main optical processes that occur in Raman spectroscopy.

These three Raman-independent processes have different effects on the detection of Raman signals in Raman spectroscopy and are eliminated or minimized via different approaches. Absorption weakens the signal considerably but a high-power laser and a high-sensitivity camera can largely solve this problem. Rayleigh scattering and anti-Stokes Raman scattering can be rejected from detection simply by using high-quality notch filters. Among these confounding factors, the most challenging one is the fluorescence since it generates red-shifted signals, which readily overlap with the spectrum of Stokes Raman scattering. One way to reduce the fluorescence in Raman spectroscopy is to switch to an excitation source at a longer wavelength, which leads,

however, to a lower Raman scattering cross-section. Reduction of fluorescence can also be accomplished by employing background removal algorithms since Raman bands are much narrower and shaper than fluorescent bands (63).

2.2.2   Raman micro- and macro-spectroscopy

A schematic of a typical laboratory Raman system is shown in Figure 2.2. A monochromatic laser with high average power is required and directed into either an optical microscope (micro-spectroscopy, or confocal Raman microscopy) or focused onto a macroscopic sample (macro-spectroscopy, or regular Raman spectroscopy). Backscattered light is collected by the objective and then sent to a spectrometer following appropriate filtering. The excitation source can be at various visible and near-infrared wavelengths. Longer-wavelength excitations are usually preferred when examining biological samples since they produce less fluorescent background.
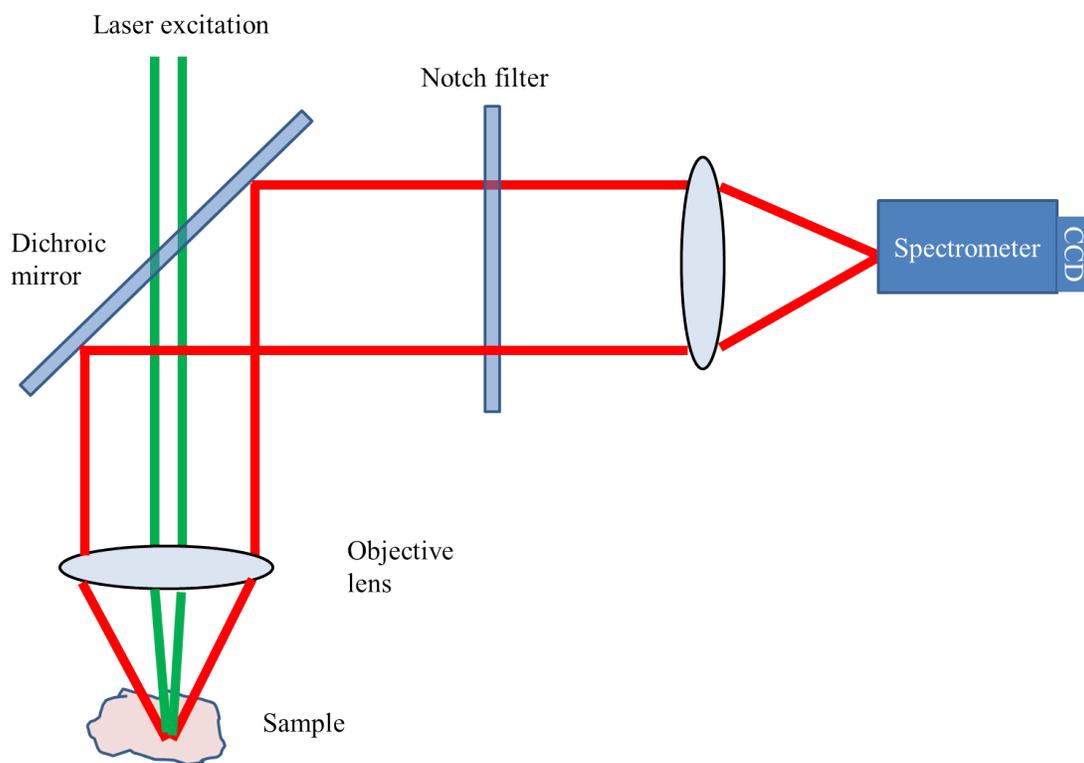

Figure 2.2 Schematic of Raman spectroscopy.

The main difference between Raman macro-spectroscopy and micro-spectroscopy is that the latter has a much smaller probed volume and therefore a higher spatial resolution. This approach avoids the volume averaging effect, but at the sacrifice of SNR, working distance, and field-of-view. Whether the researcher should choose Raman macro-spectroscopy or micro-spectroscopy depends highly on the application. For clinical investigations, in which examination speed and area are a priority, Raman macro-spectroscopy is preferred as it has much more Raman-active molecules within the probed volume, thereby resulting in better SNR and larger field-of-view. A good example is that, via an optical fiber probe, Raman spectroscopy can provide real-time, even intraoperative, detection of brain cancer (64). Nevertheless, identifying and understanding the biochemical basis underlying these potential applications is an important and necessary step toward developing more powerful and reliable diagnostic tools. Furthermore, when the biochemical changes are subtle and only occur at the microscopic level, they can rarely be detected by Raman macro-spectroscopy. In these circumstances, Raman micro-spectroscopy is commonly used, for its higher spatial resolution and chemical specificity.

2.3   Raman spectroscopy: application

Raman spectroscopy provides quantitative, label-free, and molecule-specific analysis of all major chemical constituents of the sample. Despite the rich information contained in the Raman spectrum, the development of Raman spectroscopy for further applications has been largely inhibited due to the extremely weak Raman signals. However, during the past few decades, there has been rapid developments in instrumentation, including new lasers, filters, spectrometers, and cameras. Raman spectroscopy has been extensively utilized in biomedical and clinical investigations, for instance, for delineating cell and tissue types based upon compositional changes

in minerals and proteins (65), tracking changes in the abundance of mRNA and proteins throughout stem cell differentiation (66), and discriminating types and grades of neoplasms (67).

Cancer continues to persist as one of the leading causes of death throughout the world. Early detection and improved diagnosis are the key to efficacious cancer treatments. Since Raman spectroscopy has the capability to provide detailed chemical information, including the biochemical changes accompanying carcinogenesis, it has been widely used for the diagnosis of a wide variety of cancers. For instance, much interest has been devoted to the detection of skin cancer, the disease that one in five Americans will develop in the course of a lifetime (68). A real-time Raman spectroscopy has been demonstrated for the *in vivo* diagnosis of skin cancer with sensitivities between 95% and 99% and specificities ranging between 15% and 54% (69). Another important application is the detection of invasive brain cancer cells, which cannot be achieved using current clinical technologies. For this purpose, a recent study reported a handheld Raman spectroscopy probe that can be used for the intraoperative detection of cancer cells in the human brain with a sensitivity of 93% and a specificity of 91% (64).

A large number of studies have investigated the utilization of Raman spectroscopy for the diagnosis of breast cancer. For instance, in 2005, Raman spectroscopy was demonstrated to be able to distinguish malignant tissues from normal and benign tissues with 94% sensitivity and 96% specificity (70). The spectral differences mainly come from the significant increase of collagen content and the decrease of fat content in the cancerous tissues. Two years later, it was reported that low, intermediate, and high nuclear-grade DCIS (ductal carcinoma *in situ*) and Grade I-III IDC (invasive ductal carcinoma) can be differentiated based on a series of Raman bands defined collectively from previous studies (71). This study showed that Raman spectroscopy has a strong potential role for guiding decisions and for the staging of breast cancer.

As mentioned above, the power of Raman spectroscopy has been demonstrated in the diagnosis of various cancer types by analyzing the abundance of chemical components within the tumor (64,70,72). However, its ability to characterize chemical changes outside the tumor, such as in the tumor micro- and macroenvironment, has been underappreciated. As Raman signals from one single point arise from all chemical components present within the probed volume, a considerable part of the rich information carried within the signal can be easily compromised by the complex mixtures of chemicals, as well as by the noise and background from the instrument and sample. Thus, it is proposed that if these effects are minimized, there is a chance that one could identify previously unseen changes in the tumor micro- and macroenvironment.

The objective of this study, therefore, is to characterize the alteration of PUFA abundance in the tumor micro- and macroenvironment in human breast and rat mammary cancer by utilizing the power of Raman micro-spectroscopy. Raman micro-spectroscopy was chosen for this study to avoid the volume averaging effect and to ensure optimal spatial resolution and chemical specificity (73).

2.4 Statement of work

This thesis presents Raman micro-spectroscopic analysis of fatty acids in the tumor and the tumor micro- and macroenvironments in human breast and rat mammary cancer. In Chapter 3, techniques and algorithms employed in the data acquisition and processing are introduced. Close attention is paid to the acquirement and extraction of high-SNR, fatty-acid dominant spectra that allow for the detection of the subtle but reproducible biochemical changes in the tumor micro- and macroenvironments. Chapter 4 presents the application of this spectroscopic analysis for characterizing tumor and the tumor micro- and macroenvironments in rat mammary cancer. In Chapter 5, the spectroscopic analysis is applied to characterizing malignant and normal-appearing

tissue in human breast cancer. The increased PUFA levels identified at the tumor sites, and within the tumor micro- and macroenvironments in both rat models and human subjects demonstrate the power of this spectroscopic analysis, and may provide new insights into the mechanisms of how breast cancer regulates its environment. A summary of the presented work and suggestions for future work are discussed in Chapter 6.

# 3 Data Acquisition and Processing

3.1 Data acquisition

3.1.1 Experimental setup

The experimental setup of the confocal Raman microscope is shown in Figure 3.1. A continuous-wave diode-pumped solid-state laser (Ventus, Laser Quantum) operating at a 532 nm center wavelength is used as an excitation source in a commercial confocal Raman microscope system (Horiba LabRAM HR 3D, Horiba). The beam is directed into a 0.25 N.A. 10× objective, resulting in a laser spot size of 2.6 μm at the focus. The beam out of the objective is focused directly onto the tissue with a power of ~ 20 mW. Back-scattered light is collected by the same objective and sent to a spectrometer. The collected spectral range is 500 to 3500 cm^{-1} at a spectral resolution of 3 cm^{-1}. Wavenumbers are calibrated with spectroscopic features from pure oleic acid by linear fitting. A bright-field microscope is used to focus and choose regions of interest. In this particular study, areas that appear to be dominated by lipid droplets, as shown in the bright-field image Figure 3.2 are more likely to be selected for the analysis of fatty acids.

Because SNR is crucial for identifying the subtle alterations in the tumor micro- and macroenvironment, 532 nm was chosen as the excitation wavelength, as the Raman intensity is inversely proportional to the fourth power of the laser wavelength (Eqn. (2.10)). The main disadvantage associated with using shorter wavelengths is the presence of strong autofluorescence in the Raman spectra. Nevertheless, it was found that the autofluorescence from the sample can largely be quenched by 500-ms illumination with focused excitation light at each point prior to data acquisition. In addition to the choice of laser, a 10x objective was chosen over other objectives with larger magnification number because it has a relatively longer working distance and is more likely to have an acceptable SNR throughout the scanning process of the sample.
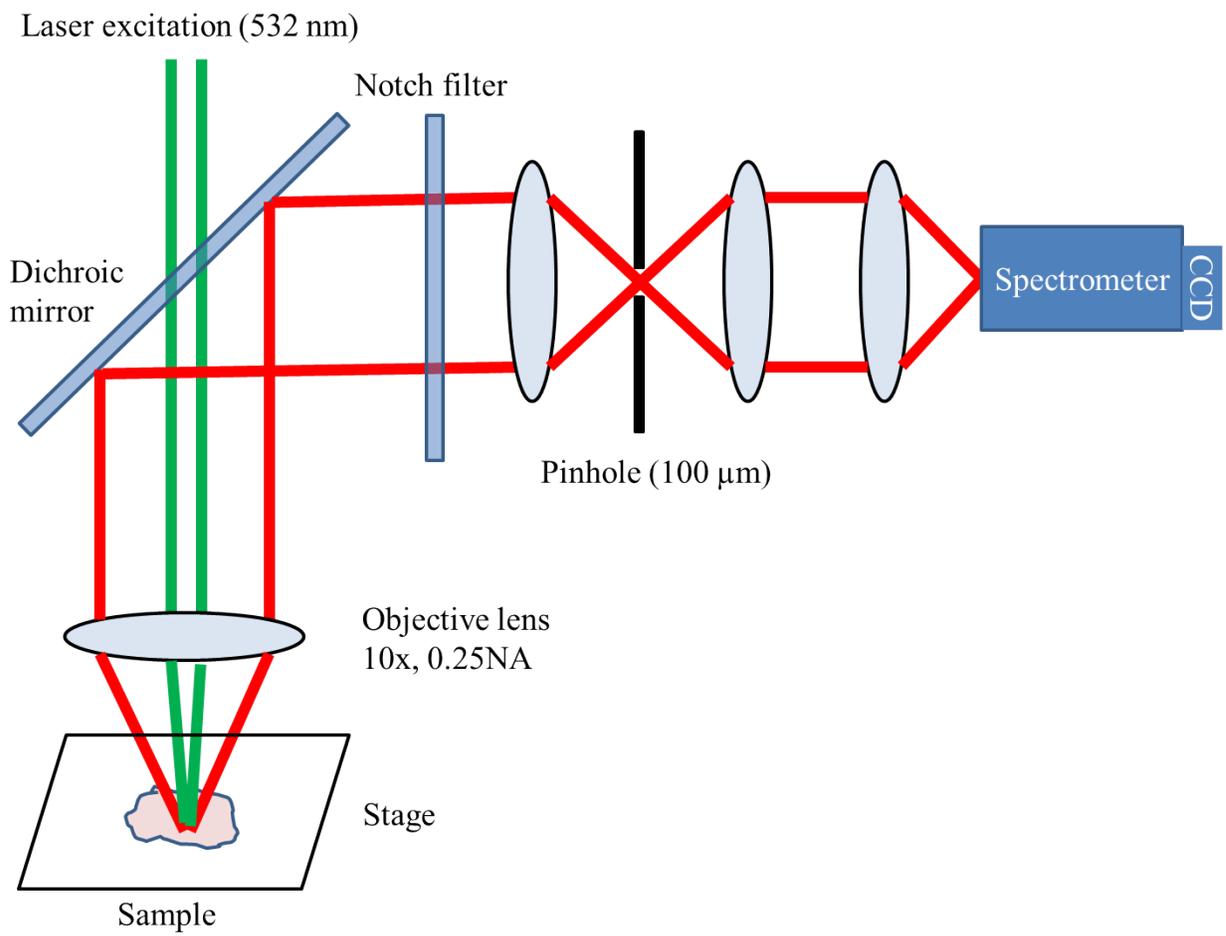
Figure 3.1 Experimental schematic of the confocal Raman microscope.
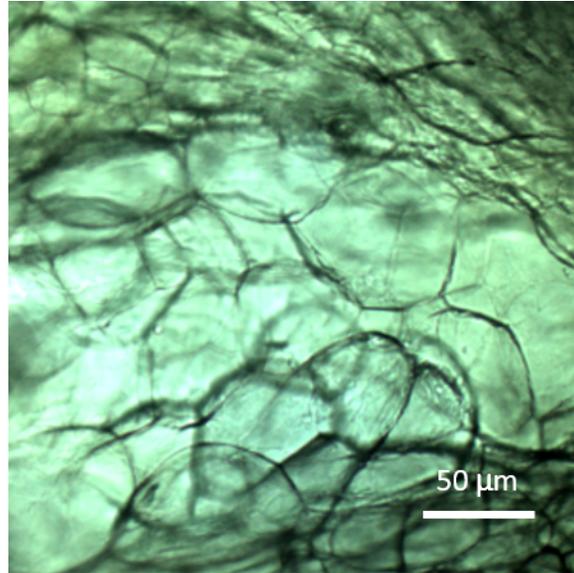
**Figure 3.2 Bright-field image of a tissue area dominated by lipid droplets in fresh human breast tissue.**

3.1.2 Parameters for data acquisition

The parameters for data acquisition were chosen with the intention to minimize the noise and volume averaging effect in this study. Raman spectroscopy measurements were performed in a two-dimensional grid with a step size of 100 µm and a grid size that ranged from 10×10 to 20×20. Scanning was performed using a motorized stage. The acquisition time for each point was 6 s. Although 0.5-1.0 s was sufficient to obtain a discernable spectrum, a longer acquisition time was preferred to reduce noise. Throughout the imaging session, the tissue sample was kept hydrated with drops of a saline solution. The selected areas varied from 1 mm^2 (100 spectra) to 4 mm^2 (900 spectra), depending on the size and the shape of the tissue sample. These acquisition times were consistent with state-of-the-art commercial confocal Raman spectroscopy systems (52) .

3.2 Data Preprocessing

3.2.1 Wavenumber calibration

Even on the same instrument, day-to-day drift in the laser wavelength or the alignment of the spectrograph may result in offset errors ranging from 0 to 30 wavenumbers. Hence, wavenumber calibration was required every day to ensure the reproducibility of the interpretation of the spectra. The wavenumbers were calibrated with spectroscopic features from pure oleic acid (Sigma-Aldrich) by linear fitting, as shown in Figure 3.3.
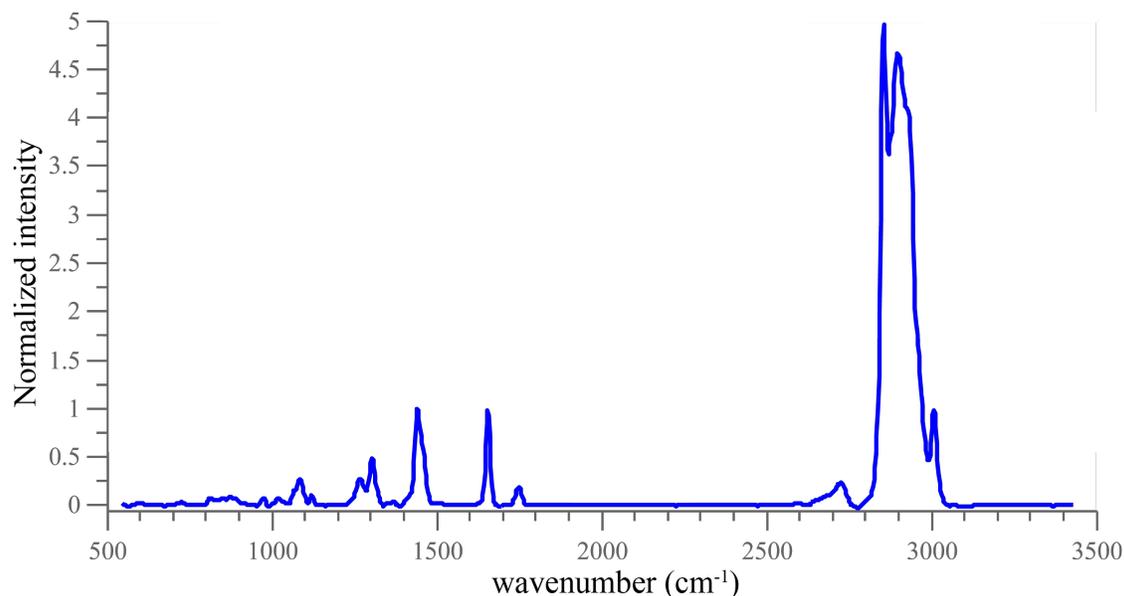
Figure 3.3 Raman spectrum of pure oleic acid.

3.2.2 SNR filtering

Although the SNR was the priority during data acquisition, there was still a chance that some data points would have low-quality spectra due to an uneven tissue surface or sample contamination. The inclusion of these highly noisy data could easily obfuscate the subtle spectral differences between the fatty acid subtypes. As a result, the filtering of high-SNR spectra was performed prior to any further data analysis.

It was found that spectra of low SNR were usually accompanied by these features: large fluctuations, saturation, and especially strong signals in the silent region (1800-2700 cm^{-1}). Thus,

spectra with low SNR could be effectively removed by identifying these features. Specifically, if any band in the silent region of a spectrum was greater than 1/5th of the highest peak throughout the entire Raman spectrum (500-3500 cm^{-1}), then this spectrum was removed. Similarly, if any peak was identical to its neighbors, the corresponding spectrum was classified as being a saturated signal and removed from the data pool.

3.2.3   Fluorescence removal

It was expected that the Raman spectra of biological samples would often accompanied by observable fluorescence at the excitation wavelength of 532 nm, even after fluorescence quenching. Fortunately, the spectra from fluorescence have much more slowing varying peaks than Raman spectra. Based on this signature, they could be removed by a variety of methods, including band pass filtering in the Fourier domain, point difference derivatives, Savitsky-Golay derivatives, and baseline fitting. In this study, fluorescence was removed by baseline fitting and correction (msbackadj; MATLAB; Mathworks, Natick, Mass.). Figure 3.4 shows the spectrum before and after baseline correction.
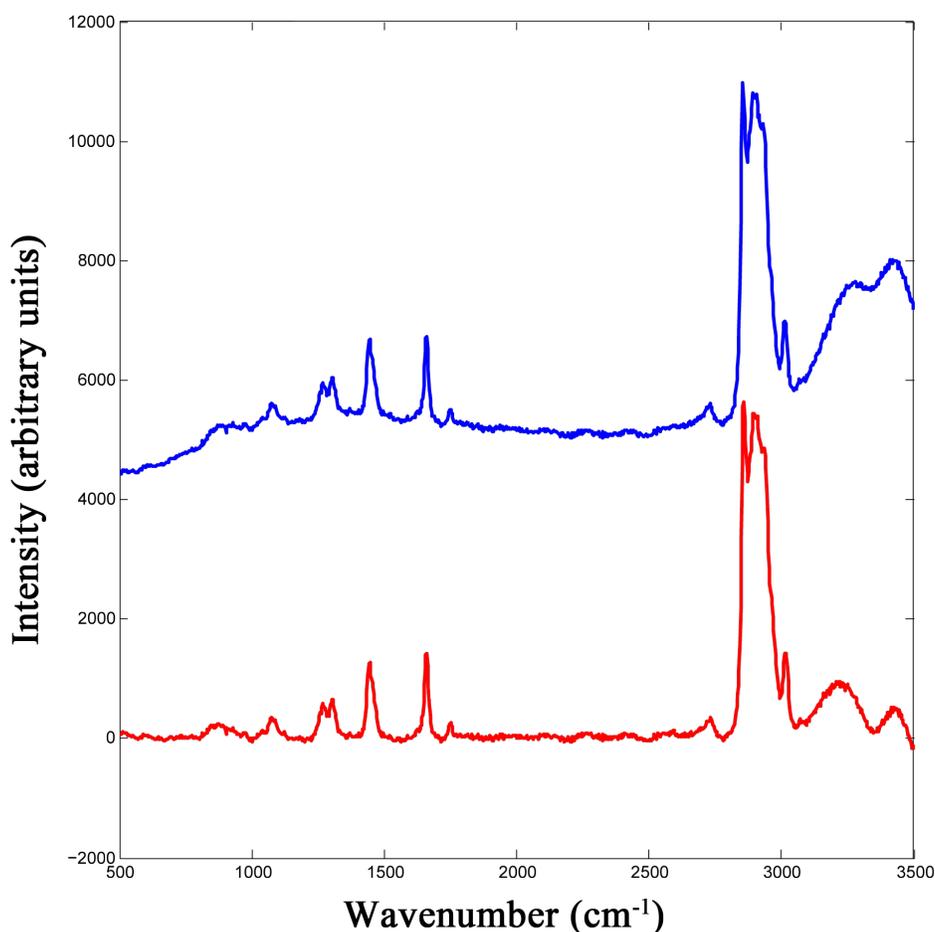
Figure 3.4 Fluorescence removal. Curve in blue is the Raman spectrum before baseline correction. Curve in red is the Raman spectrum after baseline correction.

3.2.4 Normalization

According to Beer's law, spontaneous Raman scattering has a linear dependence on the analyte concentration (74). However, the day-to-day fluctuations of the instrument made it impossible to apply Beer's law to the absolute Raman intensity. Therefore, normalization or ratiometrics was a requirement for reliable quantification of Raman signals. In the present study, Raman spectra were normalized against the peak at 1440 cm^{-1} (CH$_2$ bending mode of proteins and lipids (75)).

25

3.3 Extraction of unsaturated fatty acids

3.3.1 Pattern recognition algorithm

A pattern recognition algorithm was employed to match the acquired Raman spectra against a pre-defined library of fatty acid (FA) spectra so that interference from other chemical species would be minimized.

The extraction of fatty acids can be accomplished by either decomposition methods or classification methods. These two methods hold different assumptions of the dataset and inevitably introduce different types of artifacts into the final results. For instance, multiple linear regression (MLR), a commonly used decomposition method, works well for the decomposition of a spectrum with a limited number of distinct basis elements. However, given the fact that there are at least nine basis spectra present in breast tissue and the orthogonality between these basis spectra is not ideal (70,76), the risk of overfitting by decomposition will be high. The artifacts resulting from this overfitting problem will vary from sample to sample and can strongly inhibit the identification of the subtle spectral differences between normal and normal appearing tissue. Therefore, decomposition is not suitable for this application and classification methods were explored. Normally this option would result in an undesired rejection of a large proportion of the data which are not classified as fatty acids. However, in this particular study, the rejection rate is not an issue. The first reason is that fatty acids are abundant in breast tissues (except in solid tumors). Second, during data acquisition, areas dominated with lipid droplets (Figure 3.2) were purposely preferred, thereby increasing the number of fatty-acid dominated samples. Finally, fatty acids have an extremely strong Raman scattering cross-section compared to the other main chemical components of breast tissue (76). Therefore, classification is a reasonable approach for this particular study.

Principle component analysis (PCA) has been shown to be very effective in spectroscopic analysis where the classification of benign and malignant tissue is needed (77). However, it needs to be noted that the computed principal components from PCA are not necessarily correlated to the studied property. In this particular study, the correlation between the studied property and the classification results needs to be strong since the ultimate goal of this study is not only to differentiate normal appearing breast tissue from healthy tissue, but more importantly, to discover the specific biochemical changes in the tumor micro- and macroenvironment in breast cancer. Under this circumstance, it is important to build a classifier based on the chemical constituents of the tissue sample.

The design of a classifier depends highly on the availability of labeled samples. If the respective sample sizes of each class are large and can be assumed to be representative of the actual probability distribution, Bayes-decision based methods, such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and Naïve Bayes classifier, can be used as they optimize the expected misclassification rate and produce optimal results (78). However, in this study, the size of the training data is limited. Thus, our classifier has to rely on the distance or similarity in feature space, which leaves k nearest neighbor (kNN), support vector machine (SVM), logistic regression, and decision trees. Furthermore, one must consider the nature of the Raman spectral content, which is a mixture of all chemical constituents present in the probed volume. The Euclidean distance metric used in kNN provides a straightforward way to filter out those spectra that are not dominated by fatty acids. Therefore, kNN was the best choice for this study and scenario. A more detailed description of the kNN approach will be given in Section 3.3.3.

3.3.2 Library design

As the extraction of fatty acids relies on matching tissue spectra to a pre-defined library of fatty acids, the construction of this pre-defined library of pure chemical compounds is an important and necessary step. The library was composed of reference spectra obtained from pure chemical compounds that are the main components of rat mammary and human breast tissue. These main components include unsaturated fatty acids (oleic acid and linoleic acid), the cell nucleus (DNA), cell cytoplasm (actin), collagen (collagen type I, III, IV and V), and cholesterol (76). These compounds were purchased from Sigma-Aldrich and were used without modifications. The basic spectra for the construction of the library are shown in Figure 3.5. Saturated fatty acids were excluded in the analysis because it was found that saturated fatty acids account for very few of the spectroscopic features observed. In addition, Raman spectra from saturated fatty acids and spectra from unsaturated fatty acids were not sufficiently similar to be represented by one group.
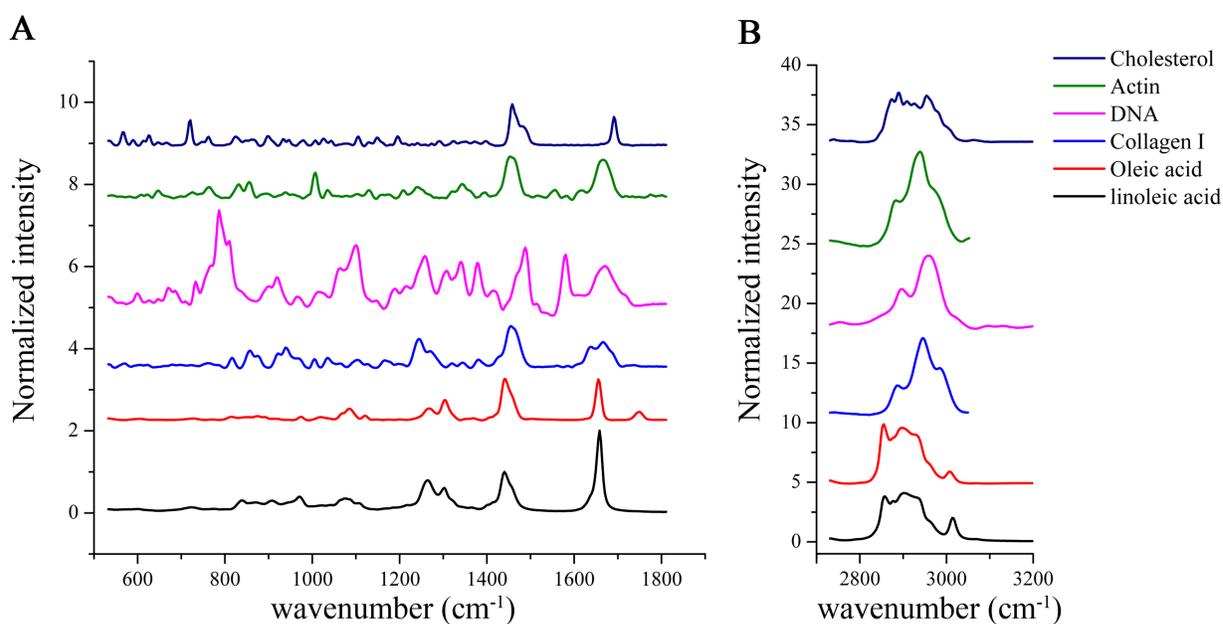
Figure 3.5 Raman spectra library of the six pure chemicals that are the main contributors to Raman signals from mammary tissue. (A) Fingerprint regime. (B) CH regime. All the spectra were preprocessed and normalized to the intensity at 1440 cm^{-1}.

3.3.3   kNN classification

The kNN classification is a simple yet effective non-parametric classification method. It consists of three main steps. (1) Searching k training samples that are nearest to the input x. The set of these k training samples are denoted as $N_k(x)$. (2) Counting the number of samples in each class in $N_k(x)$. (3) Estimating the probability of x belonging to class c by the following empirical fraction:

$$p(y = c|x, K) = \frac{1}{K} \sum_{i \in N_k(x)} \text{II}(y_i = c).$$

(3.1)

Intuitively, every input sample finds the k most similar samples in the training data, then these k training samples vote for the class of the input data based on their own classes. The similarity between samples is usually measured by Euclidean distance. An example for the Euclidean metric in two-dimensional feature space is given by:

$$d(x, x') = \| x - x' \|_2 = \sqrt{\sum_{i=1}^{d} (x_i - x'_i)^2}.$$

(3.2)

Thus, the k most similar samples are also the k nearest neighbors of the input data.

Although the idea behind this method is quiet simple, in theory when there exists enough training samples, the kNN classifier can reach within a factor of 2 of the best possible performance (79). Additionally, in practice, as demonstrated previously in the MNIST database of handwritten digits, the kNN classifier can achieve performance comparable to the convolutional neural network with proper settings (test error rate: 1.1% vs 0.7%) (80).

It should be noted that the performance of kNN methods could easily suffer from the curse of dimensionality. In the case of spectrum classification, if one were to use the raw spectrum as the input vector, the input vector will have more than 1000 elements, and so will the dimension of the feature space in the kNN methods. The Euclidean distance of two-dimensional space is shown

in Eqn. (3.2). When it is expanded to a d-dimensional space, the Euclidean ball of radius r centered at x is described by:

$$B(x,r) = \{x' \in R^d : \| x - x' \| \leq r\}. \tag{3.3}$$

Its volume is denoted by:

$$Vol[B(x,r)] = \frac{(\pi r^2)^{d/2}}{\Gamma(d/2+1)} \tag{3.4}$$

in which Γ is the Gamma function. For $d = 2k$, this results in:

$$\Gamma(d/2+1) = k!. \tag{3.5}$$

Applying Sterling's approximation, this becomes:

$$k! \quad \sqrt{2\pi} k^{k+\frac{1}{2}} e^{-k} \tag{3.6}$$

when k approaches positive infinity. Inserting Eqn. (3.5) and Eqn. (3.6) into Eqn. (3.4) gives:

$$\log Vol[B(x,r)] = k \log(\pi r^2) - \log(k!) \tag{3.7}$$
$$\sim -(k+1/2)\log(k) + k \log(\pi r^2 e) - \log(\sqrt{2\pi})$$
$$\sim -\infty.$$

Hence, as the dimension $d = 2k$ approaches positive infinity, the volume of a sphere with finite radius always tends to be zero. This phenomenon makes it very difficult to determine the closest neighbors in high-dimensional space. Therefore, instead of using the raw spectra, a subset of the spectral features was used for classification in order to reduce the dimension of the feature space.

A total of 14 spectral features were selected based on their uniqueness to each chemical component and their impact on the final classification: 787, 816, 855, 936, 1001, 1031, 1127, 1243, 1303, 1312, 1342, 1581, 1655, and 1745 cm^{-1}. The assignments for these Raman bands can be found in Table 3.1. After feature selection, the intensity I_{w_i} at the wavelength w_i was used as one of the features of the spectrum. More specifically, each spectrum was represented by a feature

vector $[I_{w_1}, I_{w_2}, \ldots, I_{w_n}]$. The kNN classification was then applied to assign a class label to the input spectrum based on the basis spectra represented by the k closest neighbors of the feature space. Furthermore, two modifications were made to the kNN classifier to make it more suitable for this particular study. (1) After the classification decision was made, to ensure the dominance of fatty acids in every spectrum, every input data x that has a distance larger than a preset distance l to the nearest training sample in the corresponding class was classified as "mixture". (2) An iterative scheme was used to generate training data based on the pure chemical samples.

Matching results for fatty acids in the rat mammary tissue samples are shown in Figure 3.6. The high degree of similarity between the reference fatty acid spectra and the observed tissue spectra, along with the narrow standard deviation of the extracted tissue spectra, demonstrates the efficiency of the matching algorithm.

One modification was made to the matching algorithm for human studies due to the presence of strong resonant signals from β-carotene found in human breast tissue but not in the rat mammary tissue. Since β-carotene is strong in the fingerprint vibration region but almost silent in the CH vibration region, and fatty acids have unique spectroscopic features in the CH region compared with other major basic components, fatty acids acquired from human breast tissue were classified based on the CH vibration region.
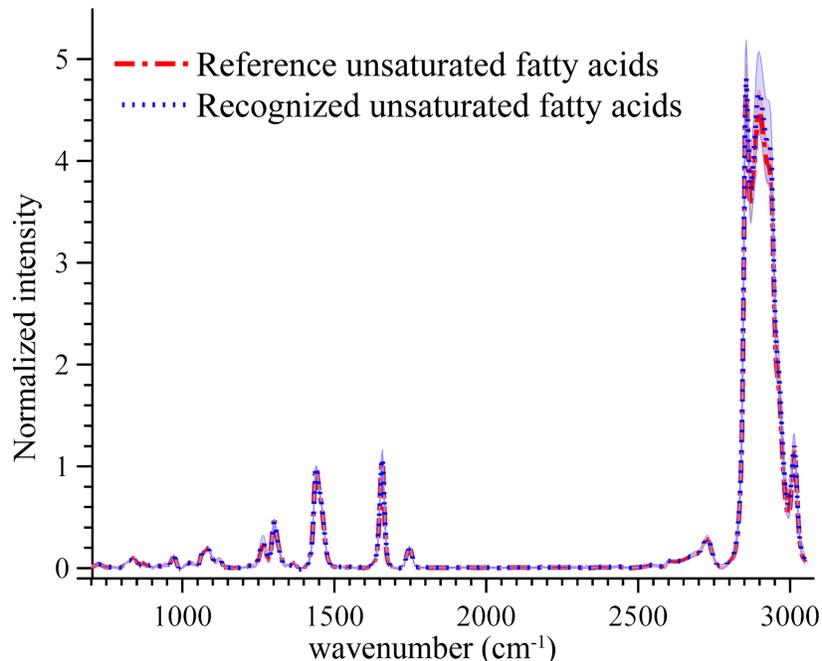
Figure 3.6 Unsaturated fatty acids were extracted with high accuracy by using a matching algorithm. The library of unsaturated fatty acids includes multiple spectra from pure oleic acid and linoleic acid, which most of the unsaturated fatty acids resemble in their Raman spectra. The standard deviation is represented by the shaded area. The small standard deviation demonstrates the high precision of our classification methods for unsaturated fatty acids.

3.3.4 Decomposition of spectrum by ordinary least square regression

Previous studies showed that β-carotene is extremely lipophilic and often found in conjunction with fat throughout the human breast (76). To avoid possible signal interference between β-carotene and fatty acids, multiple linear regression was employed prior to peak analysis to retrieve the relative concentration of fatty acids and β-carotene within a single spectrum. Multiple linear regression is an extension of the simple linear regression model; it attempts to retrieve the relationship between multiple predictor variables and one response variable by fitting a linear model to the observed data. The general multiple linear regression model takes on the form:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i$$

(3.8)

where y_i is the real response for the i th observation, x_{ij} is the j th predictor of the i th observation, β_0 is the regression intercept, and β_j is the regression slope for the j th predictor, which is the relative concentration of the j th chemical component in this application. In order to simplify the optimization problem, Eqn. (3.8) can be written in matrix notation as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\varepsilon} . \tag{3.9}$$

To estimate the best model parameters, ordinary least squares (OLS) was used to minimize the sum of the squared loss function SS_e. This OLS problem can be written as:

$$\min(SS_e) = \min_{\beta} \| \mathbf{y} - \mathbf{X\beta} \|^2 . \tag{3.10}$$

Taking the first derivative of the squared loss function with respect to $\mathbf{\beta}$ gives the OLS solution:

$$\hat{\mathbf{\beta}} = (\mathbf{X^T X})^{-1} \mathbf{X^T y} . \tag{3.11}$$

Thus, with an accurate model given by $\hat{\mathbf{\beta}}$, decomposition of spectra into fatty acid and β-carotene can be performed, allowing the fatty acid content to be extracted from the total Raman intensity. Model fits to a sample Raman spectrum are shown in Figure 3.7. The flatness of the residual spectra demonstrates that accuracy of the designed model.

In this particular model, the basis spectra for the model were acquired from pure β-carotene, oleic acid, and linoleic acid. Because oleic acid and linoleic acid have a very high degree of similarity, they were not sufficiently orthogonal to be easily differentiated in the multiple linear regression. In addition, to avoid overfitting, it was important to involve as few elements as possible while retaining important spectroscopic features. Hence, it was reasonable to use one basis element to represent both monounsaturated fatty acids and PUFAs in the model by extracting their common spectroscopic feature. Therefore, instead of using three full Raman spectra, two sets of several Raman peaks were used to implement model fitting: 1005 cm^{-1} (carotenoids), 1154 cm^{-1} (β-

carotene), 1515 cm^{-1} (β-carotene), 1440 cm^{-1} (unsaturated fatty acids), 2303 cm^{-1} (β-carotene), 2661 cm^{-1} (β-carotene), 2897 cm^{-1} (unsaturated fatty acids). More detailed assignment information for these Raman bands can be found in Table 3.1.
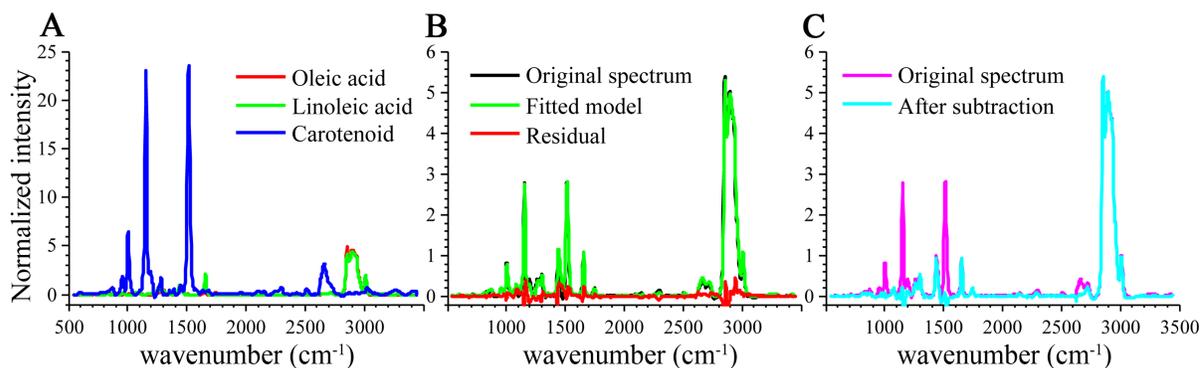
Figure 3.7 Removal of β-carotene from spectra by model fitting. (A) Raman spectra of three pure chemical components involved in the model. (B) Results of model fitting. Common peak signatures of the two fatty acids were used as one element and the unique peak signatures of carotenoid were used as the other basic element. Model fitting of the entire spectrum is not used to avoid overfitting due to similarity between the two unsaturated fatty acids. (C) Based on the model fitting coefficients, the fatty acid content is extracted by subtracting the portion of β-carotene.

3.4 Raman peak analysis

After the matching process, spectra that were dominated by features of fatty acids were selected for the peak analysis. In the peak analysis, Gaussian-peak fitting was used to decompose overlapping-peak spectrum into its component parts. Peak intensity was then calculated by finding the local peak of the fitted Gaussian line. All spectral analyses were performed in MATLAB.

3.4.1 Peak fitting

For independent peaks, peak intensity can be easily acquired by finding the local maximum. However, when peaks overlap with each other, peak analysis by simply finding local maximum at the peak of interest introduces undesired contribution from the neighbor peaks. This describes the situation for one of the Raman peaks that this study relied on, the Raman band at 3010 cm^{-1}, which lies at the tail of the neighbor peak around 2900 cm^{-1}, as shown in Figure 3.8. Decomposition of

34

the overlapping-peak spectrum into its individual components must be performed prior to further statistical analysis.

There are several considerations that must be made in order to make a peak fitting model scientifically reliable and meaningful. Assumptions about the vibrational line shapes have to be made. The observed overall line shape is a sum product of all the individual vibrations, which depends on the environment of each individual molecule. For instance, in solids, where the molecules are locked in space and the excited molecule relaxes before suffering from incoherence, the resulting line shape is Gaussian. On the other hand, in gases, where rapid motions occur frequently and loss of coherence is the dominant source of energy loss, the line shape takes on a Lorentzian profile. In liquids, where the molecules are not fixed in space but not as mobile as those in gases, the line shape is a combination Gaussian-Lorentzian profile. In the case of breast tissue, a Gaussian profile fits the Raman line shape the best and produces the smallest error among all three profiles.

Furthermore, it is important to have certain prior knowledge about the biochemical meaning of individual peaks in order to make the fitting procedure scientifically meaningful. First, fitting with a large number of peak components usually produces a nearly perfect residual, but is useless for interpretation and prone to be overfitting. Thus, predicting the number of major peak components based on the chemical structure and the acquired line profile is needed prior to peak fitting. Similarly, an initial guess of the position, width, and height of the individual peaks could help improve the accuracy and optimization speed of the peak fitting procedure. Given these considerations, peak fitting was performed with the following parameters as shown in Figure 3.8.
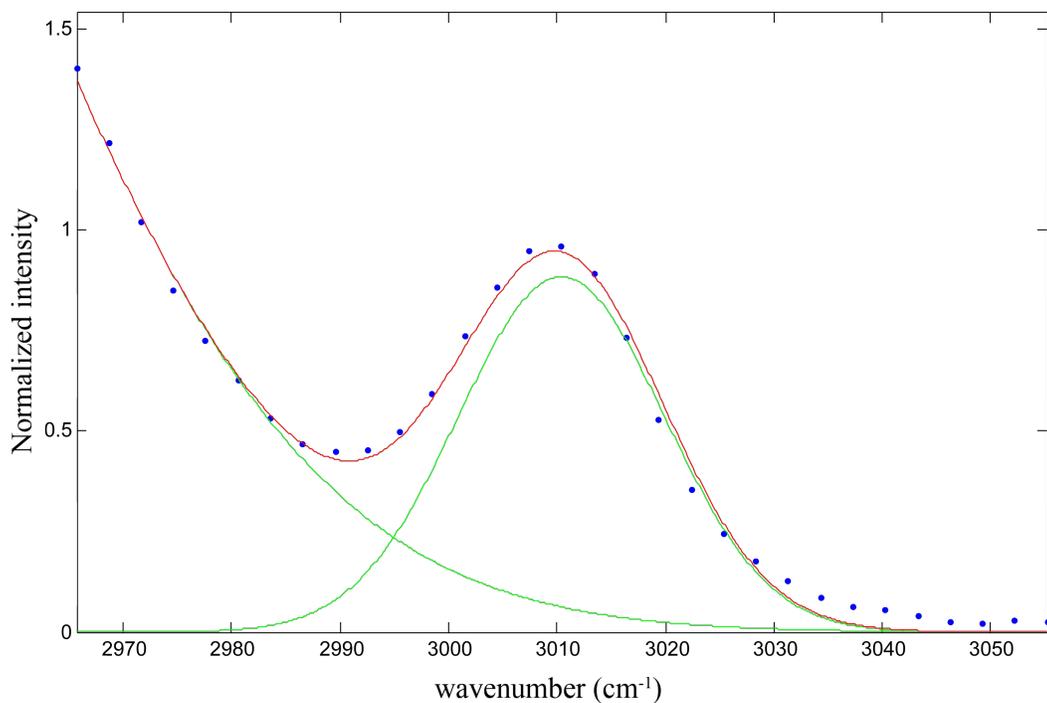
Figure 3.8 Peak fitting. The blue dots represent the original data points of the Raman spectrum. The two green curves are the resulting basis peaks via decomposition. The red curve is the fitted line based on the fitting results.

3.4.2   Raman band assignment

Table 3.1 provides the band assignments for the Raman bands that were used in this work.

Table 3.1. Assignments of the major Raman bands.

Wavenumber (cm^{-1})	Assignment
787	DNA Can be taken as a measure for the relative quantity of nucleic acids present
816	C-C stretching of collagen
855	Proline
936	Proline (collagen type I)
1001	Phenylalanie (collagen)
1005	β-carotene
1031	C-N stretching of proteins
1127	C-N stretching of proteins
1154	C-C (&C-N) stretching of β-carotene
1243	Collagen
1267	C=C groups in unsaturated fatty acids
1303	CH$_2$ deformation of lipids
1312	CH$_3$CH$_2$ twisting mode of collagen
1342	DNA/RNA
1440	CH$_2$ bending of fatty acids
1515	C-C stretch mode of β-carotene
1581	C=C bending mode of phenylalanine
1655	C=C stretching of lipids
1745	C=0 stretching of lipids
2303	OH-NH-CH stretching regions of β-carotene
2661	C-H stretch of β-carotene
2897	CH stretch of lipids and protein
3010	Unsaturated =CH stretch

3.4.3 Statistics

To determine whether tissue samples from healthy and cancerous subjects had significantly different degrees of fatty acid unsaturation, a Kruskal-Wallis test was used (kruskalwallis; MATLAB), followed by a multiple comparison test (multcompare; MATLAB) with α =0.05.

The Kruskal-Wallis test is a rank-based nonparametric (distribution free) method that can be used to test whether the samples of two or more groups are from identical distributions. It is sometimes also called the "nonparametric one-way ANOVA" and is usually used when the assumptions of one-way ANOVA are not met. In a one-way ANOVA, it is assumed that the

distribution of each group follows a normal distribution and the variance within the groups is approximately equal. However, in this study, as shown by the density distributions in Figure 4.2 and Figure 5.2, the data from normal appearing tissue do not take the form of a normal distribution and have much larger variation than the data from cancerous and healthy tissue. Therefore, the Kruskal-Wallis test was used in this study to test whether data from different locations of cancerous and healthy rats share the same distribution.

It should be noted that the Kruskal-Wallis test only provides information on whether there is a statistically significant difference between groups; it cannot determine where the difference lies between specific groups. To find out which specific groups are statistically different from each other, the post hoc test was used. The prerequisite for running a post hoc test is a significant Kruskal-Wallis result, which shows an overall significant difference between the groups.

4   Quantification of PUFAs in rat mammary cancer

4.1   Study design

An *ex vivo* imaging study was performed to investigate the biochemical changes in the tumor micro- and macroenvironment of mammary cancer in a pre-clinical rat model using Raman micro-spectroscopy. Fresh rat mammary tissues were obtained from a total of 10 female rats and imaged immediately after their removal.

Animal procedures were conducted in accordance with a protocol approved by the Institutional Animal Care and Use Committee at the University of Illinois at Urbana-Champaign. To induce mammary tumors in the female Wistar-Furth rats (Harlan, IN), N-nitroso-N-methylurea (NMU; 12.5 mg/mL; Sigma, St. Louis, MO) was injected intraperitoneally at a concentration of 55 mg/kg into the left side of the abdomen when the animals were 7 weeks old. One week later, the same amount of NMU was injected intraperitoneally into the right side of the abdomen. After approximately 12 weeks of age, when the mammary tumors became palpable, primary tumor sites and non-tumor mammary tissue at different distances from the primary tumors were removed and immediately assessed spectroscopically.

The data collected from rat mammary tissue were divided into four groups: tumor sites, tumor microenvironment, tumor macroenvironment, and control. Tumor sites were harvested from the solid palpable primary tumors. The tumor microenvironment refers to surrounding mammary tissue within a distance of 1 cm from the center of the primary tumor. The tumor macroenvironment represents tissues collected more than 3 cm away from the primary tumor. The control group is for healthy mammary tissue from rats with no treatment. More specific information about the number of samples can be found in Table 4.1.

**Table 4.1. Rat mammary tissue samples used in this study.**

Rat	Type	Rats (n)	Samples (n)
Experiment	Injected with NMU		
	Tumor sites	5	10
	Surrounding tissue 1 cm away from center of visible tumor	5	17
	Mammary tissue more than 3 cm away from the center of visible tumor	5	13
Control	Injected with PBS		
	Healthy mammary tissue	5	37
Total		10	77

4.2 Results and discussion

4.2.1 Spectral signatures of PUFAs

Figure 4.1 provides the spectral analysis of the fatty acids from the tumor sites, the tumor microenvironment, the tumor macroenvironment, and control samples. Consistent with previous results, Raman spectra of fatty acids in normal mammary tissue appear to be dominated by Raman features of oleic acid, a monounsaturated fatty acid that is the most abundant fatty acid in mammalian cells (81). Conversely, a higher average degree of unsaturation (3010 cm^{-1}) was found in the tissue of tumor-bearing rats, implying an increase in PUFA levels.

However, as the peak at 3010 cm^{-1} is overlapping with the tail of the strong neighboring CH band, the higher intensity observed at 3010 cm^{-1} in cancerous rats could be attributed to the stronger base line from the CH region rather than from increased PUFA levels. Decomposition of the overlapping peaks and systematic quantification were needed to provide a more objective and comprehensive analysis of fatty acid content for each type of tissue samples.
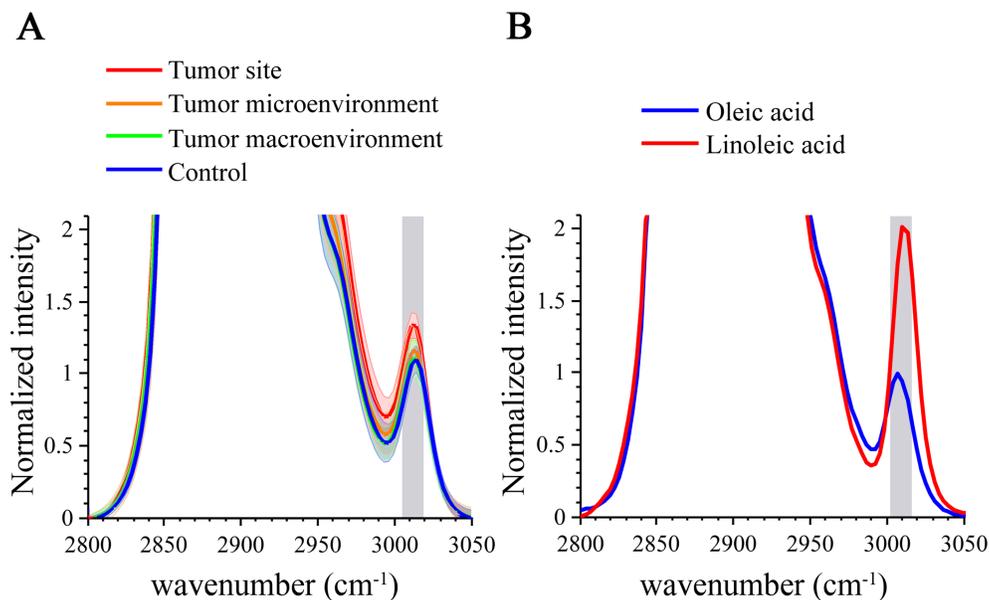
Figure 4.1 Average tissue spectra and representative reference spectra. A. comparison of average Raman spectra of fatty acids in the mammary tissue from tumor-bearing rats, and healthy mammary tissue from control rats. Shaded curves show the corresponding standard deviations of all spectra acquired from each group. The major peak used for quantifying the degree of unsaturation is marked with shaded vertical bands (3010 cm^{-1}). B. Raman spectra of two pure reference fatty acids (oleic acid and linoleic acid) in the CH region from 2800 cm^{-1} to 3050 cm^{-1}.

4.2.2 Distribution of PUFAs

For a better understanding of the trends observed in these spectra, quantification of the entire tissue sample was performed to analyze fatty acid composition. By Beer's law, the ratio of the Raman intensity at 3010 cm^{-1} ($=C-H$ stretching vibration (82)) to the intensity at 1440 cm^{-1} (CH$_2$ deformation (83)), I_{3010}/I_{1440}, depends linearly on the ratio of the number of $=C-H$ groups to the number of CH_2 groups (84). A ratio value ($N_{=C-H}/N_{CH_2}$) higher than 1/7 (oleic acid) indicates more than one double bond, thereby indicating the existence of PUFAs. After peak decomposition, I_{3010}/I_{1440} was measured from each spectrum to obtain histograms of the degrees of unsaturation of fatty acids in the four groups. The density distribution of I_{3010}/I_{1440}, which represents the degree of unsaturation of the fatty acids, is shown in Figure 4.2. Both tumor and

control groups were found to have a unimodal Gaussian-like distribution for I_{3010}/I_{1440} values, with significantly different mean values, suggesting a large-scale compositional shift from monounsaturated fatty acids to PUFAs in solid tumors. In contrast, a bimodal/trimodal distribution was observed in the tumor micro- and macro-environments. The bimodal/trimodal distribution reflects the presence of two or three distinct types of fatty acid mixtures that are well-separated in space, with one mode similar to that of the control group (mostly monounsaturated fatty acids) and the other modes similar to that of the tumor group (larger proportion of PUFAs). These results suggest that a tumor-supportive environment and healthy tissue may differ by only a relatively small subset of fatty acids.
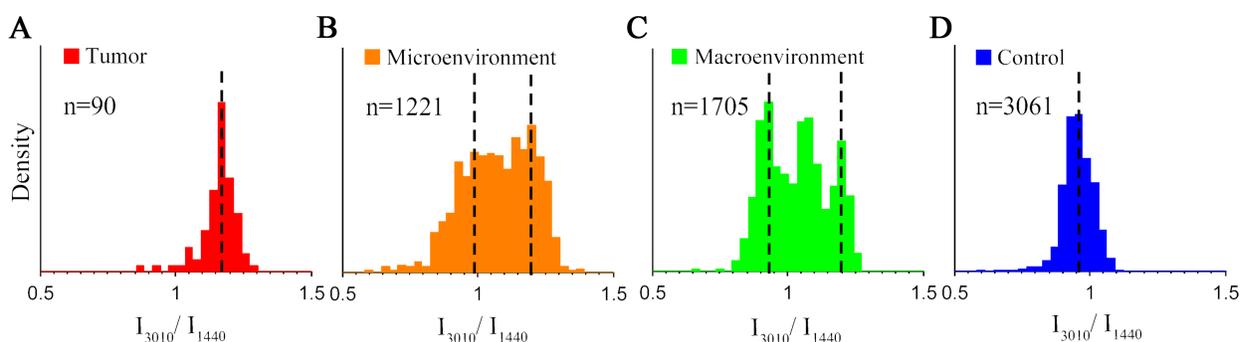
Figure 4.2 Density distributions of the level of unsaturation of fatty acids from four different groups (5 rats in each group): A. tumor, B. tumor microenvironment, C. tumor macroenvironments, and D. control (no treatment group).

4.2.3 Increased PUFAs in cancerous rats

For a more intuitive interpretation of I_{3010}/I_{1440}, a linear model for predicting the degree of unsaturation, was developed based on the measured spectra and known chemical structure from reference fatty acids: oleic acid (C18:1) and linoleic acid (C 18:2) (84,85). The average value of I_{3010}/I_{1440} was used to predict the average $N_{=C-H}/N_{CH_2}$ for each tissue sample. Figure 4.3 shows that tissue samples from tumor-bearing rats exhibited a significantly higher ratio of $N_{=C-H}/N_{CH_2}$ than those from the control group (mean ± standard deviation: 0.1548 ± 0.0035), regardless of

whether they were from the solid tumor site (0.1937, no standard deviation is associated with this measurement since only one sample was qualified for fatty acid analysis as explained below), tumor microenvironment (0.1760 ± 0.0107), or tumor macroenvironment (0.1798 ± 0.0088). It is worth noting that compared with other samples, only one solid tumor sample had a sufficiently high concentration of fatty acids for analysis because the Raman signals from the solid tumor areas tend to have greater contributions from the cell cytoplasm, cell nucleus, and collagen, but much less from fatty acids, which is consistent with what was reported in previous studies (70,75).
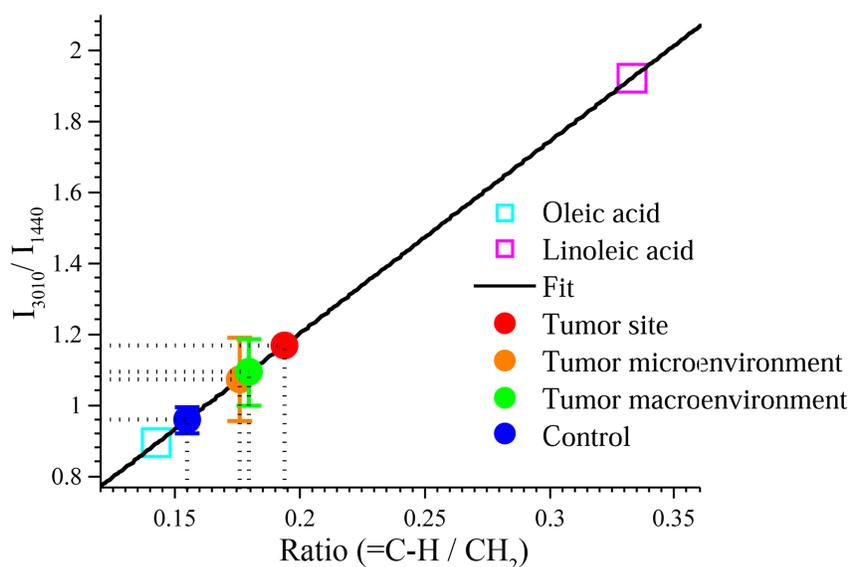

Figure 4.3 Quantitative analysis of $N_{=C-H}/N_{CH_2}$ of fatty acids in each tissue sample. The Raman measurements of the extracted fatty acids from tissue (I_{3010}/I_{1440}) were fitted on the linear model to predict the molecular structure ($N_{=C-H}/N_{CH_2}$).

4.2.4   Correlation to the density of cancer cells

Since it has been suggested that increased PUFA levels can be positively associated with the presence of metastatic cancer cells (52), the possibility that the observed alteration of fatty acid composition in the tumor macroenvironment could be due to a local influence resulting from a small number of disseminated cancer cells, instead of a global effect of cancer progression, was considered. To address this, an evaluation of the correlation between PUFA levels and the abundance of cancer cells was needed.

First, the abundance of cancer cells at different tissue sites was measured. The 787 cm^{-1} Raman peak (DNA: O-P-O, cytosine, uracil, thymine (83)), a signature for DNA, cell nuclei, and cancerous tissue, was used to quantify the relative concentration of nucleic acids present in the tissue (86). As shown in Figure 4.4A, high concentrations of DNA were only observed at the tumor sites, which was consistent with many reports of using Raman spectroscopy for tumor detection (52,86).

Next, the effect of the density of cancer cells on the concentration of PUFAs in tissue was tested. As shown in Figure 4.4B, the data points representing tissue from cancerous rats mostly lie in the upper half of the graph, which corresponds to significantly increased PUFA levels, but exhibited no consistent trend in terms of the density of cancer cells. The poor correlation between the relative DNA concentration and PUFA accumulation, together with the absence of strong DNA bands in the tumor micro- and macroenvironment, demonstrates that the observation of increased production of PUFAs is independent from the presence of densely-accumulated cancer cells and likely a realistic reflection of more global macroenvironmental changes in lipid metabolism.
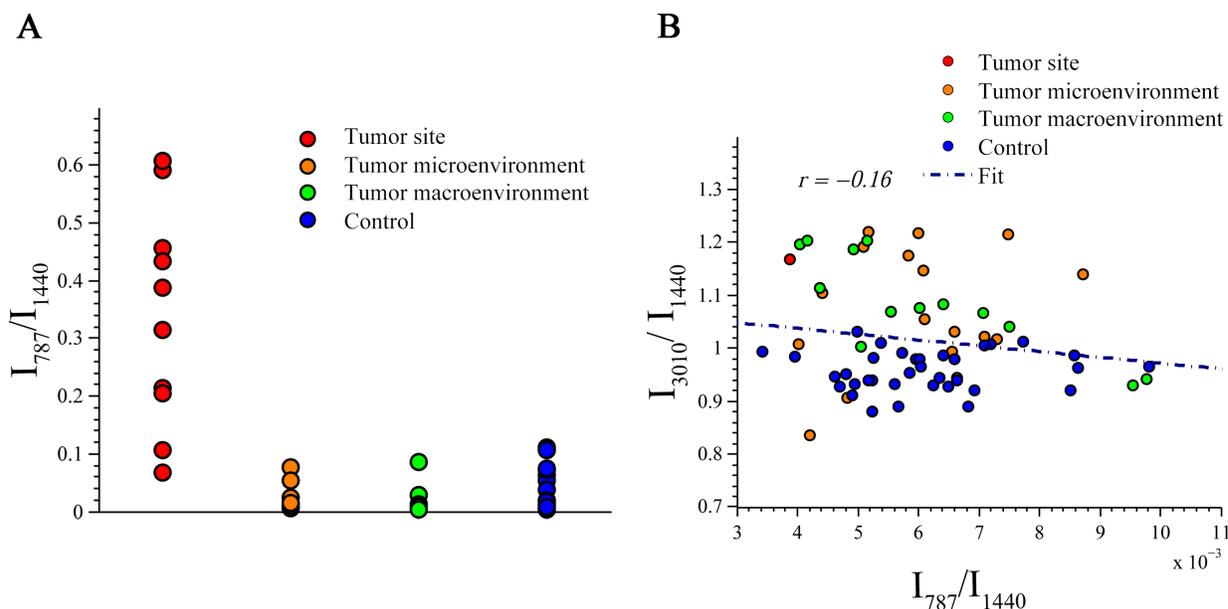
Figure 4.4 The lack of a strong DNA signal and poor correlation between DNA and PUFA levels in the tumor micro- and macroenvironment show that increased PUFA levels are not the consequence of disseminated cancer cells in the rat model. One mean DNA concentration value was generated for each specimen and grouped in terms of their location and cancer status. A. Scatterplot of relative DNA concentration of different tissue sites in rats. B. Correlation plot of relative DNA concentration and the unsaturation degree of fatty acids in rats.

4.2.5   Correlation to the fingerprint region of the Raman spectrum

As 90% of the spectroscopic Raman information on biological tissue is carried in the fingerprint stretching region (from 400 cm^{-1} to 1800 cm^{-1}), a considerable proportion of the Raman instruments have been designed for signal detection in the fingerprint region. Additionally, the strong C=C stretch at 1655 cm^{-1} and the =C-H in-plane deformation at 1267 cm^{-1} have both been shown to indicate the degree of unsaturation in the chain. Thus, it is reasonable to extend our spectral analysis from the Raman CH stretching band region (2800-3100 cm^{-1}) to the fingerprint vibration region (1200-1800 cm^{-1}). Although the Raman intensity at 1267 cm^{-1} has been useful in studies of unsaturated lipids, it is often obscured by the CH$_2$ stretching band at 1303 cm^{-1}. Thus, the C=C stretch at 1655 cm^{-1} is used as an indicator in the fingerprint region for the quantification of the degree of unsaturation of fatty acids.

Figure 4.5 shows the analysis of I_{1655}/I_{1440}, the ratio of the Raman intensity of the C=C stretching vibration to the intensity of the CH_2 deformation (84). As expected, results from the two spectral ranges in the rat study are strongly correlated and demonstrate a similar trend that the primary tumor, the tumor microenvironment, and the tumor macroenvironment have higher PUFA levels, compared to the healthy tissue environment. In previous analyses, the Raman CH vibration was chosen over the fingerprint vibration region because β-carotene is often found in conjunction with fat throughout the human breast (76), and produces strong Raman signals in the fingerprint vibration region but not in the CH vibration region (87,88) at the excitation wavelength of 532 nm.
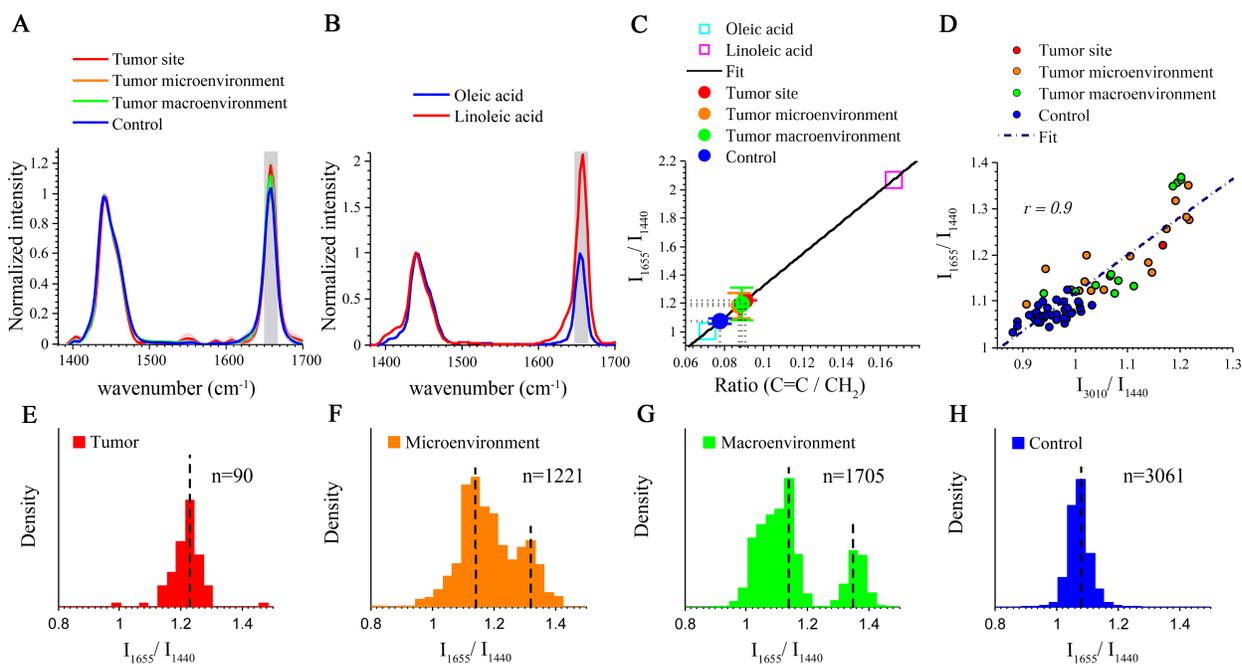
Figure 4.5 Analysis in the fingerprint region also shows increased PUFA levels in the tumor and in the micro- and macroenvironments in the rat mammary gland. A. Comparison of average Raman spectra of fatty acids from tumor sites, the micro- and macro tumor environments of tumor-bearing rats, and the healthy tissue from control rats. B. Raman spectra of two pure reference fatty acid (oleic acid and linoleic acid) in the fingerprint region from 1400 cm^{-1} to 1700 cm^{-1}. C. Quantitative analysis of the unsaturated degree of fatty acids from different tissue sites. D. Correlation plot of Raman intensity at 1665 cm^{-1} and 3010 cm^{-1} from fatty acids of various sites in cancer and control rats. E-H. Density distributions of the degree of unsaturation of fatty acids from four different groups (5 rats in each group): tumor, micro- and macroenvironments, and control (no treatment group).

4.3   Discussion

The data presented here suggests that the primary tumor, tumor microenvironment, and tumor macroenvironment from rats induced with NMU have higher PUFA levels, compared with the healthy mammary tissue from control rats. However, one should be cautious to draw the conclusion that breast cancer development causes increased PUFA levels across the rat mammary tissue.

In 2004, it was reported that the mammary epithelial cells exposed *in vitro* to NMU went through neoplastic transformation when they were injected into a carcinogen-treated stroma but

formed phenotypically normal ducts when injected into non-treated stroma (89). Thus, it was suggested that the stroma is a crucial target of NMU in rat mammary gland carcinogenesis. Due to such direct effects of NMU on the stroma, it is intuitive to question whether the fatty acid content alteration observed in the micro- and macroenvironment of NMU-treated rats was directly induced by NMU for tumor initiation instead of as a consequence of the mutational event in cancer cells. A straightforward way to exclude the effect of NMU is to explore the fatty acid composition in natural human breast cancer. Thus, to address this concern, and also to explore the clinical potential of our findings, chemical analysis of cancerous human breast tissue was performed.

5   Quantification of PUFAs in human breast cancer

5.1   Study design

This study was conducted under a protocol approved by the Institutional Review Boards at the University of Illinois at Urbana-Champaign and Carle Foundation Hospital, Urbana, IL. All tissue samples were obtained from subjects who preoperatively signed an informed consent permitting the investigational use of their tissue. Fresh samples of human breast tissue were obtained from Carle Foundation Hospital and the National Development and Research Institutes, Inc. (NDRI). Human breast tissue obtained from Carle was acquired following either mastectomy or breast reduction surgery, transported in a cold saline solution, and examined *ex vivo* within 12 hours of excision. Human breast tissue obtained from NDRI was acquired from breast reduction surgery, transported in dry ice, and examined *ex vivo* within 48 hours of excision. Tissue were acquired from a total of 7 female breast cancer subjects, ages 46 through 83, with an average age of 60 years, and 6 healthy female subjects, ages 24 through 54, with an average age of 40 years. The body mass index (BMI) for cancer patients ranged from 21.2 to 35.4, with an average BMI of 25.7. For the healthy subjects, the BMI was only available from two of the subjects due to procedural problems. The BMI for these two patients was 25 and 38, respectively. More specific information about the tissue samples can be found in Table 5.1.

The data collected from human breast tissue were divided into three groups: cancerous, normal appearing, and normal. Cancerous breast tissue samples were obtained from subjects undergoing mastectomy and diagnosed by a board-certified pathologist as containing both primary tumor and non-tumor tissue that bordered the tumor. Normal appearing breast tissue samples from breast cancer subjects were obtained from areas that were at least 5 centimeters away from the

removed primary tumor and were identified as normal by pathologists. Normal breast tissue samples were obtained from subjects undergoing breast reduction surgery.

Table 5.1. Human breast tissue samples used in this study.

Human	Type	Patients (n)	Samples (n)
With Breast Cancer	Breast cancer surgery		
	Cancerous tissue from breast cancer patients	7	52
	Normal-appearing tissue from breast cancer patients	7	29
Without Breast Cancer	Breast reduction surgery		
	Normal tissue from cancer-free patients	6	38
Total		13	119

5.2 Results and discussion

5.2.1 Spectral signatures of PUFA

Figure 5.1 shows that the average intensities of the unsaturation band (3010 cm^{-1}) of the fatty acids in the tissue samples from the cancer group appear to be stronger than those in the healthy group, indicating possibly increased PUFA levels.
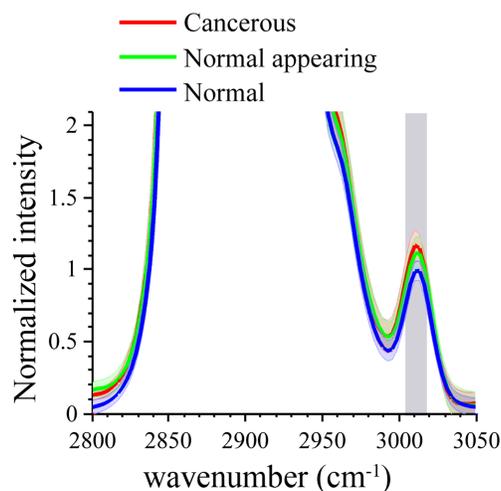
Figure 5.1 Comparison of average Raman spectra of fatty acids from cancerous and normal-appearing tissue of cancer subjects, and from normal tissue of cancer-free subjects.

5.2.2   Distribution of PUFAs

Figure 5.2 shows that, consistent with the pre-clinical rat study, both cancerous and healthy tissue samples have a unimodal distribution of the degree of unsaturation with different mean values, suggesting a large-scale transition from monounsaturated fatty acids to PUFAs at the tumor sites and nearby tumor boundaries. In contrast, the group of normal appearing tissue exhibits a bimodal pattern in the distribution of the degree of unsaturation of fatty acids, which suggests the alteration of a relatively small subset of fatty acids in the tumor macroenvironment of human breast cancer.
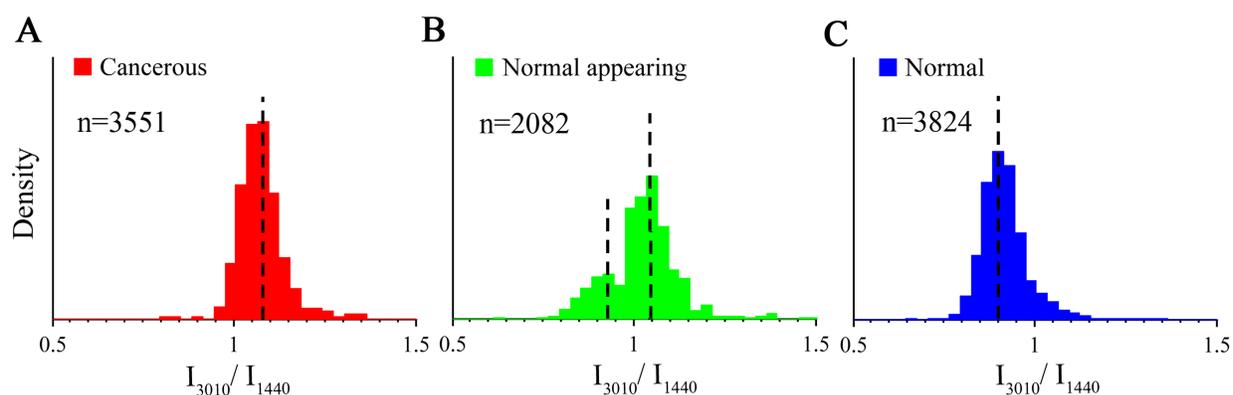
Figure 5.2 Density distributions of the degree of unsaturation of fatty acids from three groups of tissue samples: cancerous, normal-appearing, normal.

5.2.3   Increased PUFAs in cancer subjects

The next step was to investigate if significantly increased PUFA levels accompany human breast cancer. As shown in Figure 5.3, the average value of $N_{=C-H}/N_{CH_2}$ of fatty acids from healthy breast tissue is closest to that of oleic acid (0.1478 ± 0.0048 vs. 0.1429), which further confirms the predominance of monounsaturated fatty acids in healthy human breast tissue (88). Consistent with the rat study, cancerous tissues showed a higher value of $N_{=C-H}/N_{CH_2}$ compared with normal tissue from cancer-free subjects (0.1768 ± 0.0035 and 0.1478 ± 0.0048, respectively). This observation is in line with previous experimental studies reporting that PUFAs have tumor- and metastasis-promoting effects in cultured cells, animals, and human (53,54,56). Furthermore, the normal appearing tissues from cancer subjects were found to have significantly higher PUFA levels (0.1650 ± 0.0058) compared with the normal tissues from cancer-free subjects, suggesting that the regulation and alteration of lipid metabolism by cancer cells affects not only the primary tumor site and its local microenvironment, but also areas in the macroenvironment, even a few centimeters away from the primary tumor.
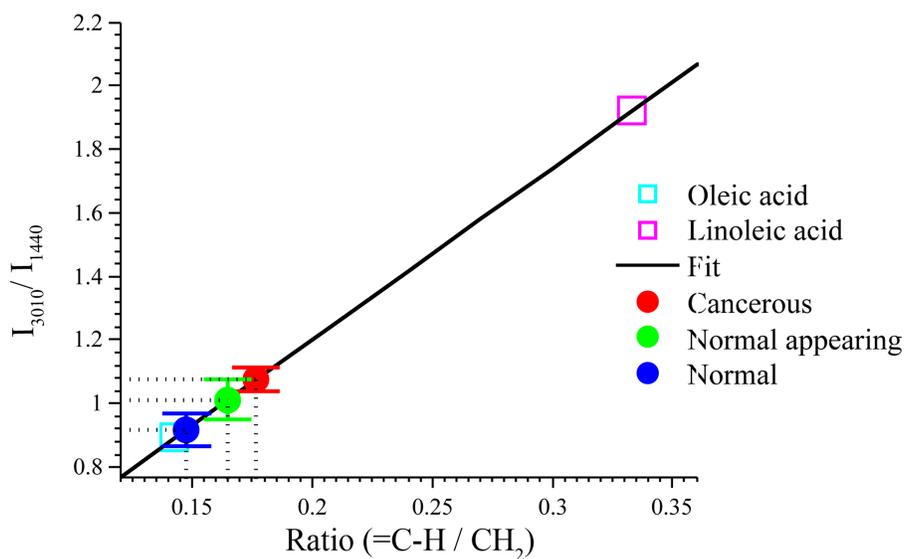
Figure 5.3 Quantitative analysis of $N_{=C-H}/N_{CH_2}$ of fatty acids in each tissue sample.

5.2.4 Correlation to the density of cancer cells

In order to evaluate the effect of the density of cancer cells on PUFA levels in human breast tissue, abundance of cancer cells was measured through the Raman peak at 787 cm^{-1} and correlations between DNA and PUFA levels were tested based on relevant Raman bands. Consistent with the pre-clinical study, high concentrations of DNA were only observed at the tumor sites, and no strong correlations existed between the DNA abundance and PUFA levels, as shown in Figure 5.4. Therefore, the possibility of disseminated cancer cells causing increased PUFA levels in human breast cancer can likely be ruled out, lending support to the previous hypothesis that increased PUFA levels is a part of the macroenvironmental regulation in breast cancer.
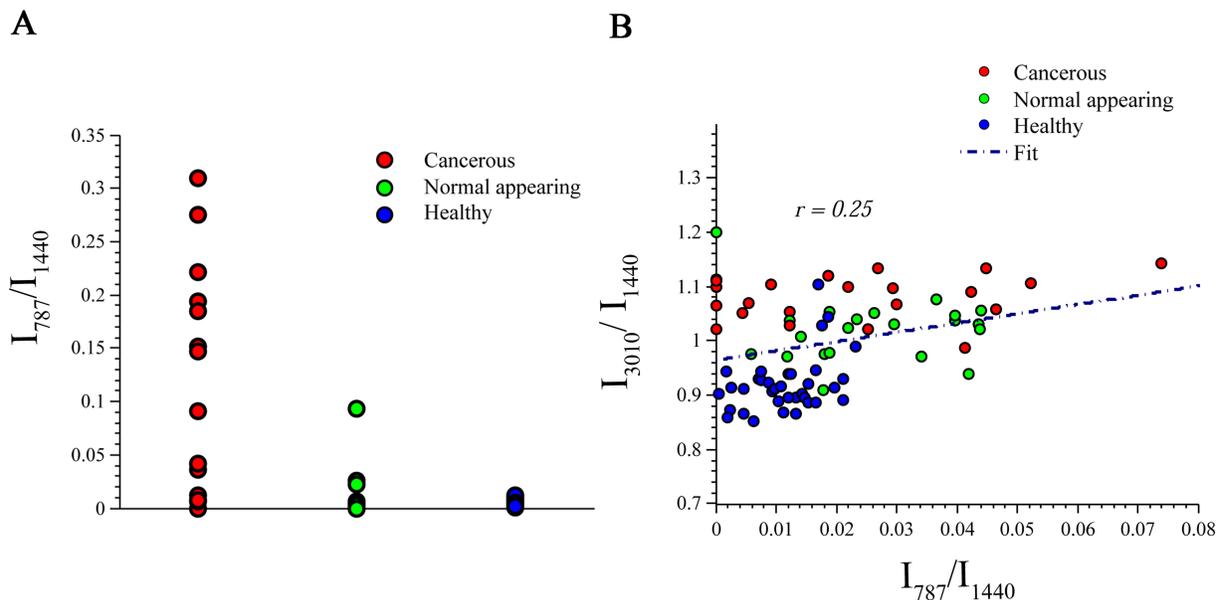
Figure 5.4 The lack of a strong DNA signal and poor correlation between DNA and PUFA levels in the normal appearing tissue show that increased PUFA levels are not the consequence of disseminated cancer cells in human breast tissue. One mean DNA concentration value was generated for each specimen and grouped in terms of their location and cancer status. A. Scatterplot of relative DNA concentration of different tissue sites in human. B. Correlation plot of relative DNA concentration and the unsaturation degree of fatty acids in human.

5.2.5   Correlation to obesity

Obesity can induce a series of systemic metabolic syndromes, including insulin resistance, elevated levels of cholesterol, and promoted inflammation. Although it is a well-known fact that obese people have a higher cancer risk and worse prognoses, the complex mechanism underlying this obesity-cancer link is poorly understood. A major challenge has been in deciphering which of these obesity-induced factors mediates the key risk factors for breast cancer.

One of the key processes that appear to play a causal role in the obesity-cancer link is inflammation. Overweight and obese women showed elevated systemic inflammation in their breast tissue (90,91). This could likely be attributed to the increased level of COX-2, which catalyzes the synthesis of proinflammatory PGE2 in overweight and obese women (90). There are several interesting observations about PGE2. First, although it is counterintuitive, PEG2, the by-

product of obesity, has been shown to be able to induce COX-2 and thereby increase its own production by itself (35). Next, PEG2 is known to promote breast cancer development by stimulating the synthesis of estrogen, inhibiting apoptosis, and facilitating migration and invasion. Finally, PGE2 is one of the n-6 PUFA derived messengers (35). Given all these facts and implications, it is intuitive to hypothesize that obesity is positively associated with elevated levels of PUFAs, which likely play an essential role in mediating the synthesis of PGE2, and thereby promoting breast cancer.

As a preliminary experiment, BMI was used as an indicator for obesity and I_{3010}/I_{1440} as an indicator for PUFA levels to assess the correlation between these two factors. A positive relationship was expected between the PUFA levels in the healthy breast tissue and the BMI of that subject. Unfortunately, no conclusion could be drawn from the preliminary results since BMI information was only available from two healthy subjects, as shown in Figure 5.5.

It should be noted, however, that BMI, a parameter based on a person's body weight and height, is not necessarily enough information to scientifically characterize their health risks from obesity. It is a little shocking to realize that this parameter has been used for the past two centuries since its conception in 1823. Nevertheless, in recent years, there has been considerable debate surrounding the use of BMI in defining obesity. In future obesity-cancer studies, it will be desirable to include the assessment of other metabolic markers such as fasting blood glucose, total serum cholesterol, and circulating lipid levels, in addition to BMI, to better characterize the health risks from obesity.
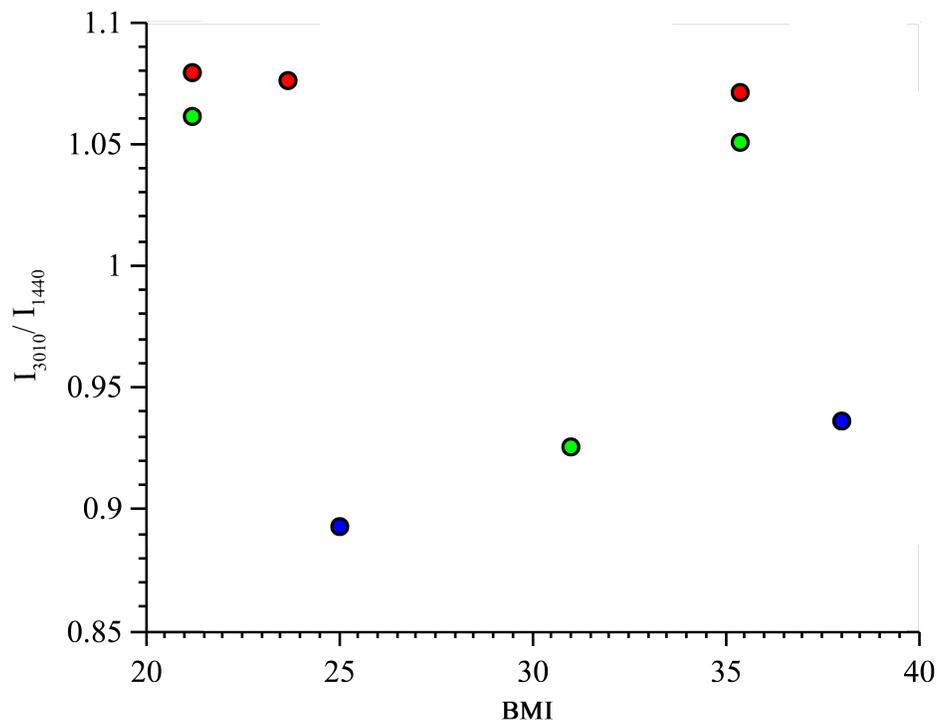
Figure 5.5 Scatterplot of PUFA level versus BMI. Each dot represents the average PUFA level and BMI of one subject. The color of each dot corresponds to which group they belong, red for cancerous tissue, green for normal appearing tissue, and blue for healthy tissue.

5.3 Discussion

The increases of PUFA levels in the normal appearing tissue from human subjects and pre-clinical rat models are in line with reports that have shown the cancer-promoting effects of PUFAs in the macroenvironment (24). It is of note that different PUFA families have different effects on breast cancer. Indeed, n-6 PUFAs have been shown to promote tumorigenesis via increasing circulating estrogenic compounds levels (47), cell proliferation rate (48), and inflammation (49), while n-3 PUFAs have been shown to be responsible for suppressing tumor growth by decreasing circulating estrogenic compound levels (47), breast cancer cell growth (50), and HER2 expression (51). These studies support the findings in this thesis research and suggest that the majority of the PUFA content investigated here are n-6 PUFAs (52). Nevertheless, further studies are needed to confirm this.

More specifically, a significantly increased abundance of n-6 linoleic acids at the tumor sites has been documented in clinical studies on patients with breast cancer (92). However, the increased PUFA levels observed in the tumor micro- and macroenvironment are unexpected and even appear contradictory to previous studies that have suggested the variation in fatty acid content between the tumor macroenvironment and healthy tissue is negligible (93). According to the distributions shown in Figure 4.2 and Figure 5.2, the tumor micro- and macroenvironment and healthy tissue only differ by a relatively small subset of fatty acids, compared with the large-scale transition of fatty acid content at the tumor sites. Differentiation by analytical methods used in the past may have been hindered by the volume averaging effect and relatively low SNR since speed and portability are usually prioritized in clinical studies.

Furthermore, studies on the links between lipid metabolism and breast cancer have been largely limited to murine models, mainly because cells involved in adipogenic regulatory networks are difficult to obtain and culture for long periods of time (94). Lipid metabolism has been shown to differ greatly between human and murine models so one must be cautious to extend the observations reported here in murine models to humans (95). Nevertheless, the consistency between the pre-clinical rat study and the human study suggests that the NMU-induced rat mammary tumor model could be used as a future pre-clinical model for studies involving PUFA production in the tumor sites, the microenvironment, and the macroenvironment.

The data presented in this thesis suggests that breast cancer is not just a localized event centered on carcinoma cells in the primary tumor, but rather a more global process involving tissue surrounding and even distant from the primary tumor. Our observation of increased PUFA levels in the macroenvironment of breast cancer, together with previous reports that have shown the positive association between PUFAs and obesity (96), raises the intriguing possibility that

increased PUFA levels in the tumor macroenvironment might be one of the biochemical links between breast cancer and obesity.

6   Conclusion and future direction

6.1   Conclusion

Altered lipid metabolism has been extensively investigated in cancer cells (52,56), but not in their micro- or macroenvironment. This imbalance of attention is due in part to the belief that normal appearing tissue distant from tumor sites could be assumed to be similar to healthy tissue and be used as a control group in comparison studies. However, the results reported here suggest that subtle but reproducible differences exist in the fatty acid content between healthy tissue from cancer-free subjects and normal appearing tissue from cancerous subjects. By optimizing high SNR and minimizing the effect of volume averaging of chemicals during Raman data acquisition and processing, significantly increased PUFA levels were identified in the tumor, as well as in the tumor micro- and macroenvironments in both human tissues and in pre-clinical rat models. These data support the concept of micro- and macroenvironmental regulation in mammary tissues of humans and rat models with breast carcinoma (20,21,24). These findings are likely to provide new insights into the mechanisms of breast cancer development and also suggest potential approaches for the assessment of the macroenvironment in human tissues and rat models.

In addition, this study demonstrates the capability of Raman micro-spectroscopy to identify the subtle but reproducible chemical changes in the tumor micro- and macroenvironment in human tissues and rat models with breast cancer. Considering the rapid advances in the development of portable Raman spectroscopy (64,97) and coherent Raman imaging microscopy (98,99), the proposed strategy presented here may help future applications related to intraoperative detection, prognosis, and assessment of macroenvironment-targeted therapies. Moreover, as a typical Raman spectrometer offers the best performance in either the fingerprint or CH region, the high correlation

demonstrated in the fatty acid analysis between these regions gives researchers the freedom to choose either vibration band of interest.

A major limitation of this study is the small number of animals and human tissue specimens involved in the statistical analysis. However, it is important to emphasize that the main purpose of this study was to propose a new hypothesis in macroenvironmental regulation and an approach for detection. To establish the clinical significance of these initial findings, chemical analysis of a larger number of samples is needed. Another potential future direction is to correlate PUFA abundance with the stage of breast cancer and the percentage of body fat (or BMI) to better understand their effects on PUFA production.

In conclusion, this thesis research demonstrated increased PUFA levels in the tumor, as well as in the tumor micro- and macroenvironment of human breast cancer and rat mammary cancer, which has important implications for further clarifying the global effect of cancer progression. These findings also suggest that there may be high potential for using Raman scattering to characterize the tumor micro- and macroenvironment of breast cancer in both pre-clinical animal studies and human clinical applications. This could furthermore influence current therapeutic approaches for breast cancer and related metabolic diseases.

6.2   Future direction

Future work can take many directions in the field of analyzing micro- and macroenvironmental alterations in cancer. One of the most important future directions is to develop an improved imaging system that has the capability to acquire the compositional, spatial, functional, and even temporal information about all major chemical components of tumor and its environment simultaneously. In theory, most of such information can be provided by spontaneous Raman scattering. However, practically, the intrinsically low signals levels of spontaneous Raman

techniques prevent their adoption as popular imaging modalities. One possible solution to this problem is coherent Raman imaging (CRI), which provides the same information about molecular vibrations but with signal levels that are orders-of-magnitude larger than spontaneous Raman scattering (100,101). In addition to CRI, other modalities, including multiphoton fluorescence, and second and third harmonic generation, will be incorporated into the same system to provide complimentary information about structural or functional changes in biological systems. In contrast to CRI, these modalities rely on contrast generated from only one or several specific structures or molecules. Despite this limit of contrast, these multiphoton techniques are currently much more sensitive and specific than CRI as long as the targeted molecule is active for the applied modality.

The results presented in this thesis encourage further exploration of PUFA levels in the tumor and the tumor micro- and macroenvironment in breast cancer. To begin, it is desirable to involve more disease types and larger sample sizes in future studies. It will be interesting to investigate if there is any correlation between the stage of breast cancer and PUFA levels in the normal appearing breast tissue. Additionally, if possible, dietary PUFA input should be tracked and even controlled for every subject in future studies, as it has been shown that fatty acid composition of adipose tissue depends heavily on dietary input (36). Furthermore, future studies should consider the correlation between PUFA abundance and obesity. This could provide insights into the mechanisms linking breast cancer and obesity.

References

1. Desantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics , 2013. CA Cancer J Clin. 2014;64:52–62.

2. Warburg O. Injuring of respiration the origin of cancer cells. Science. 1956;123:309–14.

3. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. Oncogene. 2008;27:5904–12.

4. Voeller D, Rahman L, Zajac-Kaye M. Elevated levels of thymidylate synthase linked to neoplastic transformation of mammalian cells. Cell Cycle. 2004;3:1005–7.

5. Kuhajda FP. Fatty-acid synthase and human cancer: new perspectives on its role in tumor biology. Nutrition. 2000;16:202–8.

6. Polyak K, Haviv I, Campbell IG. Co-evolution of tumor cells and their microenvironment. Trends Genet. 2009;25:30–8.

7. Walsh AJ, Cook RS, Lee JH, Arteaga CL, Skala MC. Collagen density and alignment in responsive and resistant trastuzumab-treated breast cancer xenografts. J Biomed Opt. 2015;20:026004.

8. Kalluri R, Zeisberg M. Fibroblasts in cancer. Nat Rev Cancer. 2006;6:392–401.

9. Mantovani A, Germano G, Marchesi F, Locatelli M, Biswas SK. Cancer-promoting tumor-associated macrophages: New vistas and open questions. Eur J Immunol. 2011;41:2522–5.

10. Nieman KM, Romero IL, Van Houten B, Lengyel E. Adipose tissue and adipocytes support tumorigenesis and metastasis. Biochim Biophys Acta - Mol Cell Biol Lipids. 2013;1831:1533–41.

11. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell. 2012;21:309–22.

12. Liotta LA. Tumor invasion and metastases—role of the extracellular matrix. Cancer Res. 1986;46:1–7.

13. Köninger J, Giese T, di Mola FF, Wente MN, Esposito I, Bachem MG, et al. Pancreatic tumor cells influence the composition of the extracellular matrix. Biochem Biophys Res Commun. 2004;322:943–9.

14. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol. 2012;196:395–406.

15. Tlsty TD, Coussens LM. Tumor stroma and regulation of cancer development. Annu Rev Pathol Mech Dis. 2006;1:119–50.

16. Fukumura D, Jain RK. Tumor microenvironment abnormalities: Causes, consequences, and strategies to normalize. J Cell Biochem. 2007;101:937–49.

17. Bissell MJ, Radisky D. Putting tumours in context. Nat Rev Cancer. 2001;1:46–54.

18. Condeelis J, Segall J. Intravital imaging of cell movement in tumours. Nat Rev Cancer. 2003;3:921–30.

19. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. Nat Rev Cancer. 2009;9:239–52.

20. Moinfar F, Beham A, Friedrich G, Deutsch A, Hrzenjak A, Luschin G, et al. Macro-environment of breast carcinoma: frequent genetic alterations in the normal appearing skins of patients with breast cancer. Mod Pathol. 2008;21:639–46.

21. Castaño Z, Tracy K, McAllister SS. The tumor macroenvironment and systemic regulation of breast cancer progression. Int J Dev Biol. 2011;55:889–97.

22. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: The metabolic requirements of cell proliferation. Science. 2009;324:1029–33.

23. Petricoin EF, Liotta LA. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr Opin Biotechnol. 2004;15:24–30.

24. Al-Zhoughbi W, Huang J, Paramasivan GS, Till H, Pichler M, Guertl-Lackner B, et al. Tumor macroenvironment and metabolism. Semin Oncol. 2014;41:281–95.

25. Moinfar F, Man YG, Arnould L, Bratthauer GL, Ratschek M, Tavassoli FA. Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. Cancer Res. 2000;60:2562–6.

26. Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. Nat Rev Cancer. 2004;4:551–61.

27. Mulligan HD, Beck S a, Tisdale MJ. Lipid metabolism in cancer cachexia. Br J Cancer. 1992;66:57–61.

28.   Baumann J, Sevinsky C, Conklin DS. Lipid biology of breast cancer. Biochim Biophys Acta. 2013;1831:1509–17.

29.   Medes G, Thomas A, Weinhouse S. Metabolism of neoplastic tissue. IV. A study of lipid synthesis in neoplastic tissue slices *in vitro*. Cancer Res. 1953;13:27–9.

30.   Menendez JA, Lupu R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. Nat Rev Cancer. 2007;7:763–77.

31.   Danilo C, Frank PG. Cholesterol and breast cancer development. Curr Opin Pharmacol. 2012;12:677–82.

32.   van Staveren WA, Deurenberg P, Katan MB, Burema J, de Groot LC, Hoffmans MD. Validity of the fatty acid composition of subcutaneous fat tissue microbiopsies as an estimate of the long-term average fatty acid composition of the diet of separate individuals. Am J Epidemiol. 1986;123:455–63.

33.   Hietanen E, Punnonen K, Punnonen R, Auvinen O. Fatty acid composition of phospholipids and neutral lipids and lipid peroxidation in human breast cancer and lipoma tissue. Carcinogenesis. 1986;7:1965–9.

34.   Mamalakis G, Hatzis C, de Bree E, Sanidas E, Tsiftsis DD, Askoxylakis J, et al. Adipose tissue fatty acids in breast cancer patients versus healthy control women from Crete. Ann Nutr Metab. 2009;54:275–82.

35.   Schmitz G, Ecker J. The opposing effects of n−3 and n−6 fatty acids. Prog Lipid Res. 2008;47:147–55.

36.   Marszalek JR, Lodish HF. Docosahexaenoic acid, fatty acid-interacting proteins, and neuronal function: breastmilk and fish are good for you. Annu Rev Cell Dev Biol. 2005;21:633–57.

37.   Bozza PT, Bandeira-Melo C. Mechanisms of leukocyte lipid body formation and function in inflammation. Mem Inst Oswaldo Cruz. 2005;100:113–20.

38.   Bandeira-Melo C, Bozza PT, Weller PF. The cellular biology of eosinophil eicosanoid formation and function. J Allergy Clin Immunol. 2002;109:393–400.

39.   Rennison JH, Van Wagoner DR. Impact of dietary fatty acids on cardiac arrhythmogenesis. Circ Arrhythmia Electrophysiol. 2009;2:460–9.

40.   Rajamoorthi K, Petrache HI, McIntosh TJ, Brown MF. Packing and viscoelasticity of polyunsaturated omega-3 and omega-6 lipid bilayers as seen by (2)H NMR and X-ray diffraction. J Am Chem Soc. 2005;127:1576–88.

41.   Calder PC. Polyunsaturated fatty acids and inflammation. Prostaglandins Leukot Essent Fat Acids. 2006;75:197–202.

42.   Prescott SM, Stenson WF. Fish oil fix. Nat Med. 2005;11:596–8.

43.   Schoonjans K, Watanabe M, Suzuki H, Mahfoudi A, Krey G, Wahli W, et al. Induction of the acyl-coenzyme A synthetase gene by fibrates and fatty acids is mediated by a peroxisome proliferator response element in the C promoter. J. Biol. Chem. 1995;270:19269–76.

44.   Frohnert BI, Hui TY, Bernlohr DA. Identification of a functional peroxisome proliferator-responsive element in the murine fatty acid transport protein gene. J Biol Chem. 1999;274:3970–7.

45.   Clarke SD, Armstrong MK, Jump DB. Dietary polyunsaturated fats uniquely suppress rat liver fatty acid synthase and S14 mRNA content. J Nutr. 1990;120:225–31.

46.   Ntambi JM, Miyazaki M, Stoehr JP, Lan H, Kendziorski CM, Yandell BS, et al. Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. Proc Natl Acad Sci. 2002;99:11482–6.

47.   Nagata C, Iwasa S, Shiraki M, Sahashi Y, Shimizu H. Association of maternal fat and alcohol intake with maternal and umbilical hormone levels and birth weight. Cancer Sci. 2007;98:869–73.

48.   Fiorio Pla A, Grange C, Antoniotti S, Tomatis C, Merlino A, Bussolati B, et al. Arachidonic acid-induced Ca2+ entry is involved in early steps of tumor angiogenesis. Mol Cancer Res. 2008;6:535–45.

49.   Rose DP, Connolly JM, Rayburn J, Coleman M. Influence of diets containing eicosapentaenoic or docosahexaenoic acid on growth and metastasis of breast cancer cells in nude mice. J Natl Cancer Inst. 1995;87:587–92.

50.   Menéndez JA, Vázquez-Martín A, Ropero S, Colomer R, Lupu R. HER2 (erbB-2)-targeted effects of the omega-3 polyunsaturated fatty acid, alpha-linolenic acid (ALA; 18:3n-3), in breast cancer cells: the "fat features" of the "Mediterranean diet" as an "anti-HER2 cocktail". Clin Transl Oncol. 2006;8:812–20.

51.   Yee LD, Young DC, Rosol TJ, Vanbuskirk AM, Clinton SK. Dietary (n-3) polyunsaturated fatty acids inhibit HER-2/neu-induced breast cancer in mice independently of the PPARgamma ligand rosiglitazone. J Nutr. 2005;135:983–8.

52.   Hedegaard M, Krafft C, Ditzel HJ, Johansen LE, Hassing S. Discriminating isogenic cancer

cells and identifying altered unsaturated fatty acid content as associated with metastasis status using K-Means clustering and partial least squares-discriminant analysis of Raman maps. Anal Chem. 2010;82:2797–802.

53.   Maclennan M, Ma DWL. Role of dietary fatty acids in mammary gland development and breast cancer. Breast Cancer Res. 2010;12:211.

54.   Byon CH, Hardy RW, Ren C, Ponnazhagan S, Welch DR, McDonald JM, et al. Free fatty acids enhance breast cancer cell migration through plasminogen activator inhibitor-1 and SMAD4. Lab Invest. 2009;89:1221–8.

55.   Sampath H, Ntambi JM. Polyunsaturated fatty acid regulation of genes of lipid metabolism. Annu Rev Nutr. 2005;25:317–40.

56.   Le TT, Huff TB, Cheng J-X. Coherent anti-Stokes Raman scattering imaging of lipids in cancer metastasis. BMC Cancer. 2009;9:42.

57.   Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004;4:579–91.

58.   Louie SM, Roberts LS, Nomura DK. Mechanisms linking obesity and cancer. Biochim Biophys Acta - Mol Cell Biol Lipids. 2013;1831:1499–508.

59.   Smith E, Dent G. Modern Raman spectroscopy - a practical approach. Wiley. 2005.

60.   Gremlich H-U. Infrared and Raman spectroscopy. Ullmann's Encycl Ind Chem. 2000. page 177–220.

61.   Raman CV, Kariamanikkam KS. A new type of secondary radiation. Nature. 1928;121:501–2.

62.   Haka AS. Development of *in vivo* Raman spectroscopy for the diagnosis of breast cancer and intra-operative margin assessment. 2005.

63.   Choo-Smith L-P, Edwards HGM, Endtz HP, Kros JM, Heule F, Barr H, et al. Medical applications of Raman spectroscopy: from proof of principle to clinical implementation. Biopolymers. 2002;67:1–9.

64.   Jermyn M, Mok K, Mercier J, Desroches J, Pichette J, Saint-arnaud K, et al. Intraoperative brain cancer detection with Raman spectroscopy in humans. 2015;7:1–10.

65.   Gentleman E, Swain RJ, Evans ND, Boonrungsiman S, Jell G, Ball MD, et al. Comparative materials differences revealed in engineered bone as a function of cell-specific differentiation. Nat Mater. 2009;8:763–70.

66.   Chan JW, Lieu DK. Label-free biochemical characterization of stem cells using vibrational spectroscopy. J Biophotonics. 2009;2:656–68.

67.   Gniadecka M, Philipsen PA, Sigurdsson S, Wessel S, Nielsen OF, Christensen DH, et al. Melanoma Diagnosis by Raman Spectroscopy and Neural Networks: Structure Alterations in Proteins and Lipids in Intact Cancer Tissue. J Invest Dermatol. 2004;122:443–9.

68.   Thieden E, Philipsen PA, Sandby-møller J, Christian H. Sun exposure, sun protection, and vitamin D. Am Med Assoc. 2005;294:1541–3.

69.   Lui H, Zhao J, McLean D, Zeng H. Real-time Raman spectroscopy for *in vivo* skin cancer diagnosis. Cancer Res. 2012;72:2491–500.

70.   Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Diagnosing breast cancer by using Raman spectroscopy. Proc Natl Acad Sci. 2005;102:12371–6.

71.   Rehman S, Movasaghi Z, Tucker AT, Joel SP, Darr JA, Ruban A V., et al. 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;38:1345–51.

72.   Barman I, Dingari NC, Saha A, McGee S, Galindo LH, Liu W, et al. Application of Raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy. Cancer Res. 2013;73:3206–15.

73.   Everall NJ. Confocal Raman microscopy: Performance, pitfalls, and best practice. Appl Spectrosc. 2009;63:254–62.

74.   Giese RH. Light scattering by small particles and models of interplanetary matter derived from the zodiacal light. Space Sci. Rev. 1963.

75.   Rehman S, Movasaghi Z, Tucker AT, Joel SP, Darr JA, Ruban A, et al. 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;38:1345–51.

76.   Shafer-Peltier KE, Haka AS, Fitzmaurice M, Crowe J, Myles J, Dasari RR, et al. Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*. J Raman Spectrosc. 2002;33:552–63.

77.   Kumar KK, Chowdary MVP, Mathew S, Rao L, Krishna CM, Kurien J. Raman spectroscopic diagnosis of breast cancers : evaluation of models. J Raman Spectrosc. 2008;39:1276–82.

78.   Duda RO, Hart PE. Pattern Classification and Scene Analysis. Leonardo. 1973.

79.   Cover T, Hart P. Nearest neighbor pattern classification. Inf Theory, IEEE Trans. 1967;13:21–7.

80.   LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86:2278–323.

81.   Brozek-pluska B, Jablonska-gajewicz J, Kordek R, Abramczyk H. Phase transitions in oleic acid and in human breast tissue as studied by Raman spectroscopy and Raman imaging. J Med Chem. 2011;54:3386–92.

82.   Baeten V, Hourant P, Morales MT, Aparicio R. Oil and fat classification by FT-Raman spectroscopy. J Agric Food Chem. 1998;8561:2638–46.

83.   Movasaghi Z, Rehman S, Rehman I. Raman spectroscopy of biological tissues. Appl Spectrosc. 2007;42:493–541.

84.   Wu H, Volponi J V, Oliver AE, Parikh AN, Simmons BA, Singh S. *In vivo* lipidomics using single-cell Raman spectroscopy. Proc Natl Acad Sci. 2011;108:3809–14.

85.   Li-Chan E. Developments in the detection of adulteration of olive oil. Trends Food Sci Technol. 1994;5:3–11.

86.   Draga ROP, Grimbergen MCM, Vijverberg PLM, van Swol CFP, Jonges TGN, Kummer JA, et al. *In vivo* bladder cancer diagnosis by high-volume Raman spectroscopy. Anal Chem. 2010;82:5993–9.

87.   Oshima Y, Sato H, Zaghloul A, Foulks GN, Yappert MC, Borchman D. Characterization of human meibum lipid using raman spectroscopy. Curr Eye Res. 2009;34:824–35.

88.   Surmacki J, Musial J, Kordek R, Abramczyk H. Raman imaging at biological interfaces : applications in breast cancer diagnosis Raman imaging at biological interfaces : applications in breast cancer diagnosis. Mol Cancer. 2013;12:48.

89.   Maffini M V, Soto AM, Calabro JM, Ucci AA, Sonnenschein C. The stroma as a crucial target in rat mammary gland carcinogenesis. J Cell Sci. 2004;117:1495–502.

90.   Subbaramaiah K, Morris PG, Zhou XK, Morrow M, Du B, Giri D, et al. Increased levels of COX-2 and prostaglandin E2 contribute to elevated aromatase expression in inflamed breast tissue of obese women. Cancer Discov. 2012;2:356–65.

91.   Morris PG, Hudis CA, Giri D, Morrow M, Falcone DJ, Zhou XK, et al. Inflammation and

increased aromatase expression occur in the breast tissue of obese women with breast cancer. Cancer Prev Res (Phila). 2011;4:1021–9.

92. Mamalakis G, Hatzis C, de Bree E, Sanidas E, Tsiftsis DD, Askoxylakis J, et al. Adipose tissue fatty acids in breast cancer patients versus healthy control women from Crete. Ann Nutr Metab. 2009;54:275–82.

93. Frank CJ, McCreery RL, Redd DC. Raman spectroscopy of normal and diseased human breast tissues. Anal Chem. 1995;67:777–83.

94. Peirce V, Carobbio S, Vidal-Puig A. The different shades of fat. Nature. 2014;510:76–83.

95. Harms M, Seale P. Brown and beige fat: development, function and therapeutic potential. Nat Med. 2013;19:1252–63.

96. Ailhaud G, Massiera F, Weill P, Legrand P, Alessandri JM, Guesnet P. Temporal changes in dietary fats: Role of n-6 polyunsaturated fatty acids in excessive adipose tissue development and relationship to obesity. Prog Lipid Res. 2006;45:203–36.

97. Lieber CA, Majumder SK, Ellis DL, Billheimer DD, Mahadevan-Jansen A. *In vivo* nonmelanoma skin cancer diagnosis using Raman microspectroscopy. Lasers Surg Med. 2008;40:461–7.

98. Brown PO, Botstein D, Lapidot M, Pilpel Y, Mazo A, Hodgson JW, et al. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. Science. 2008;322:1857–61.

99. Chowdary PD, Jiang Z, Chaney EJ, Benalcazar WA, Marks DL, Gruebele M, et al. Molecular histopathology by spectrally reconstructed nonlinear interferometric vibrational imaging. Cancer Res. 2010;70:9562–9.

100. Pestov D, Ariunbold GO, Wang X, Murawski RK, Sautenkov V a, Sokolov A V, et al. Coherent versus incoherent Raman scattering: molecular coherence excitation and measurement. Opt Lett. 2007;32:1725–7.

101. Camp Jr CH, Cicerone MT. Chemically sensitive bioimaging with coherent Raman scattering. Nat Photonics. 2015;9:295–305.