Published in final edited form as:
Anal Chem. 2012 August 7; 84(15): 6715–6722. doi:10.1021/ac3011439.

Precision of Raman Spectroscopy Measurements in Detection of Microcalcifications in Breast Needle Biopsies

Anushree Saha[†], **Ishan Barman**[‡], **Narahara Chari Dingari**[‡], **Luis H. Galindo**[‡], **Abdus Sattar**[†], **Wendy Liu**[†,§], **Donna Plecha**[†,§], **Nina Klein**[†,§], **Ramachandra Rao Dasari**[‡], and **Maryann Fitzmaurice**[†,*]

[†]Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

[‡]Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, USA

[§]University Hospitals Case Medical Center, 11100 Euclid Avenue, Cleveland, OH 44106, USA

Abstract

Microcalcifications are an early mammographic sign of breast cancer and a target for stereotactic breast needle biopsy. We developed Raman spectroscopy decision algorithms to detect breast microcalcifications, based on fit coefficients (FC) derived by modeling tissue Raman spectra as a linear combination of the Raman spectra of 9 chemical and morphologic components of breast tissue. However, little or no information is available on the precision of such measurements and its effect on the ability of Raman spectroscopy to make predictions for breast microcalcification detection. Here we report the precision, that is, the closeness of agreement between replicate Raman spectral measurements - and the model FC derived from them - obtained *ex vivo* from fresh breast biopsies from patients undergoing stereotactic breast needle biopsy, using a compact clinical Raman system. The coefficients of variation of the model FC averaged 0.03 for normal breast tissue sites, 0.12 for breast lesions without and 0.22 for breast lesions with microcalcifications. Imprecision in the FC resulted in diagnostic discordance among replicates only for line-sitters, that is, tissue sites with FC values near the decision line or plane. The source of this imprecision and their implications for the use of Raman spectroscopy for guidance of stereotactic breast biopsies for microcalcifications are also discussed. In summary, we conclude that the precision of Raman spectroscopy measurements in breast tissue obtained using our compact clinical system is more than adequate to make accurate and repeatable predictions of microcalcifications in breast tissue using decision algorithms based on model FC. This provides strong evidence of the potential of Raman spectroscopy guidance of stereotactic breast needle biopsies for microcalcifications.

Keywords

Raman; spectroscopy; precision; breast; microcalcifications

INTRODUCTION

Microcalcifications are an early mammographic sign of breast cancer and a target for stereotactic breast needle biopsy[1]. In up to 15% of patients, microcalcifications are not retrieved during stereotactic needle biopsy resulting in non-diagnostic or false negative biopsies [2]. There are two types of microcalcifications: type I microcalcifications composed of calcium oxalate (CAO) crystals, found almost exclusively in benign duct cysts; and type

*CORRESPONDING AUTHOR: maryann.fitzmaurice@case.edu.

II microcalcifications composed of calcium hydroxyapatite (CHA) concretions, most often found in proliferative breast lesions, including invasive and in situ cancer [3]. While mammography can detect microcalcifications, it cannot distinguish type I and II microcalcifications or microcalcifications associated with benign or malignant breast lesions. Therefore, a tissue biopsy must be performed to determine whether the microcalcifications seen at mammography are associated with cancer.

There are several groups currently studying breast cancer [4–14] and microcalcifications [15–19] using Raman spectroscopy, including our group [20–27]. Our goal here is to develop Raman spectroscopy as a clinical tool to provide real-time feedback to the radiologist during stereotactic needle biopsy procedures as to whether the microcalcifications seen on mammography will be retrieved. We were the first to demonstrate the potential of Raman spectroscopy to detect and distinguish type I and II microcalcifications and to differentiate type II microcalcifications associated with benign and malignant breast lesions [20]. Recently, we developed Raman algorithms for the detection of microcalcifications in breast needle biopsy tissue cores [21]. The decision algorithms are based on fit coefficients (FC) derived by modeling tissue Raman spectra as a linear combination of the Raman spectra of 9 components of breast tissue, including: epithelial cell nuclei (ECN) and cytoplasm (ECC); fat cells (FAT); β-carotene (β-CAR), a Raman active biomolecule present in the cytoplasm of fat cells; oxy-hemoglobin (OXY-HB); stromal collagen fibers (COLL); cholesterol-like extracellular deposits (CHOL); type I (calcium oxalate (CAO)) microcalcifications and type II microcalcifications (largely calcium hydroxyapatite (CHA)) [23]. Specifically, these decision algorithms use the model FC of total calcium (CAO + CHA), fat and collagen as diagnostic parameters and have positive predictive values of 96–97% for the detection of microcalcifications in stereotactic breast biopsy tissue cores, when compared with radiography and histopathology diagnosis as the gold standard.

In developing a test for the detection or diagnosis of any disease state, it is important to assess the precision of the measurements upon which the diagnosis is based, that is, the closeness of agreement between replicate measurements on the same or similar samples, as it is a key factor in determining the repeatability, reproducibility and reliability of the diagnosis. This is an issue of concern for all optical spectroscopy techniques in development for tissue-based disease diagnosis that has been little explored to date. Raman spectroscopy is regarded by many as a precise measurement technique, primarily stemming from its exquisite chemical specificity. Our group has previously shown that the uncertainty (error) of spectroscopic measurements is proportional to the spectral overlap between the analyte of interest and the other sample constituents [28]. Therefore the high degree of Raman specificity directly manifests itself in a relatively small spectroscopic measurement uncertainty (error). Never-the-less, it is important to quantitatively assess the precision of Raman spectroscopy measurements used in diagnostic applications, as even a small degree of imprecision may affect the accuracy of disease detection or diagnosis.

The precision of Raman measurements has been reported for non-diagnostic applications [29–31]. However, in published studies of Raman spectroscopy for tissue diagnosis, precision calculations are rarely reported, although such computations are critical to comprehensively investigate its potential for clinical translation. Some recent reports of the precision of diagnostic Raman measurements include investigations in sound tooth enamel[32] and determination of glycated albumin in human serum albumin solutions [33]. To the best of our knowledge, there are no reports to date of the precision of modeling of Raman spectra and subsequent assessment of model FC-based decision algorithm performance for breast cancer diagnosis or microcalcification detection.

Here we report the precision of *ex vivo* Raman spectral measurements - and of the model FC derived from them - which are used in our decision algorithms for the detection of microcalcifications in breast biopsy tissue cores, and discuss the implications for Raman spectroscopy guidance of stereotactic breast needle biopsies for microcalcifications.

EXPERIMENTAL SECTION

Breast Biopsy Raman Measurement Protocol

This study was performed on a subset of the Raman spectroscopy data set used to develop the decision algorithms to detect microcalcifications in breast biopsy tissue [21] for which replicate Raman spectral measurements were obtained. The data set in the initial algorithm development study consisted of Raman spectra from 159 tissue sites in breast biopsy tissue cores obtained from 33 patients (all female; ages 38–79) undergoing vacuum-assisted stereotactic core needle breast biopsy procedures in the Breast Health Center at University Hospitals-Case Medical Center. All data were obtained *ex vivo* from freshly excised biopsy specimens within 30 minutes of excision. All studies were approved by the Case Cancer Institutional Review Board and the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects. Informed consent was obtained from all subjects prior to their biopsy procedures.

Replicate sets of data needed to evaluate the precision of our Raman spectroscopy measurements were obtained from 106 tissue sites from breast biopsy tissue cores from 22 of the initial 33 patients. The optimal number of replicates was determined by estimating the intraclass coefficient (ICC) of a preliminary data set from 43 tissue sites, including normal breast tissue and lesions with and without microcalcifications, details of which are provided in Supporting Information Section S-1 [34]. The ICC was >0.90 for all model FC for 4 and 5 replicates, and was optimal for 5 replicates ranging from 0.93–0.97.

Five replicate Raman spectra were obtained from each tissue site in quick succession, by the same operator, without repositioning, using standard operating procedures and automated data collection and real-time data analysis. The time interval from first to last replicate measurement was approximately 90 s. A longer time scale was precluded by the need for prompt fixation of the breast biopsy tissue, because time-sensitive, patient care-related, ancillary studies must be performed on any biopsies with a diagnosis of invasive cancer on histopathology examination.

After data collection, the tissue sites from which Raman spectra were obtained were marked with multicolored colloidal inks to uniquely identify each tissue site. The tissue was then fixed in 10% neutral buffered formalin, routinely processed, paraffin embedded, cut into tissue sections and stained with hematoxylin and eosin (H&E). The H&E stained tissue sections were examined by an experienced breast pathologist, and histopathology and spectroscopy results subsequently compared.

The 106 tissue sites were classified as follows based on their appearance on a specimen radiograph plus histopathology: 28 normal, 60 lesions with microcalcifications, and 18 lesions without microcalcifications. The histopathology diagnoses for the 78 lesions with and without microcalcifications included: fibrocystic change (FCC), fibroadenoma (FA) and ductal carcinoma in situ (DCIS). The distribution of FCC, FA and DCIS in lesions with and without microcalcifications is provided in Supporting Information Table S-1. Of the 60 lesions with microcalcifications, 59 had type II and only 1 had type I microcalcifications. Some lesions with microcalcifications contained only a single microcalcification; others contained greater than 25 microcalcifications. Individual microcalcifications averaged 203

μm in diameter(range 18–1030 μm) and 475 μm deep in the tissue cores (range 0–1650 μm), which ranged from 1075–2850 μm in maximum thickness.

Raman Instrument and Data Analysis

The Raman spectral measurements were made using a portable, compact clinical Raman spectroscopy system (instrument and optical fiber probes), previously described in detail [21], using 830 nm excitation with a tissue sampling volume of approximately 1 mm^3. Once the Raman spectra were acquired, they were fit with a previously developed breast model [23], in which the Raman tissue spectrum is considered a linear combination of the basis spectra of: 9 breast tissue constituents (ECN, ECC, FAT, CHOL, β-CAR, OXY-HB, COLL, CAO and CHA), water and 2 optical fiber probe materials (epoxy and sapphire), yielding FC that correlate with the presence or absence of spectral contributions from the basis spectra and provide information about the morphological and chemical composition of the tissue. Real time data analysis was used to help confirm data collection from tissue sites with microcalcifications. Further details of the clinical Raman spectroscopy system and spectral data analysis are provided in the Supporting Information Section S-2 [28, 35, 36].

Statistical Analysis

Mean, standard deviation (SD) and coefficient of variation (CV), or relative SD, were calculated for each model FC for each tissue site for each tissue type (MATLAB). SD measures how widely values are dispersed from the mean, and was calculated as the square root of total variance using equation 1:

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}} \quad (1)$$

where $\bar{x}$ is mean value and n is sample size. CV is a normalized measure of dispersion and was calculated as the ratio of the SD to the mean. Since the mean FC value is the denominator in the CV calculations, CV could not be calculated for tissue sites having a mean FC of zero for the model component being considered and were arbitrarily set to zero.

RESULTS

Raman Spectra

Figure 1 shows exemplar Raman spectra from normal breast tissue and breast lesions (FCC) with and without microcalcifications from the same patient biopsy. The radiographic appearance, histopathology, Raman spectrum, model fit and FC for a typical breast lesion (FCC) with microcalcifications are provided in the Supporting Information Figure S-1. No consistent trend in absolute spectral intensity (increase or decrease) was seen over time among the 5 replicates for any one tissue site, for any of the three tissue types (normal breast tissue, lesions with microcalcifications or lesions without microcalcifications). However, the signal-to-noise ratio (SNR) of the Raman spectra did vary somewhat based on the tissue type, with breast lesions with and without microcalcifications exhibiting a lower SNR (mean SNR of 38.30 dB and 38.69 dB, resp.) than normal tissue (mean SNR of 38.87 dB).

Model Fits and Fit Coefficients

Model fits of the Raman spectra were good. However, as with SNR, model fits did vary somewhat based on the tissue type, with breast lesions with and without microcalcifications exhibiting higher mean standard deviations of the residuals (0.14 and 0.06, resp.) than normal breast tissue (0.03). No consistent change in model FC over time was observed for any of the 12 FC among the 5 replicates for any one tissue site, for any of the three tissue

types (normal breast tissue, breast lesions without microcalcifications or breast lesions with microcalcifications).

Table 1 shows the mean, range and SD for 10 FC, the FC of each of the 9 Raman model breast tissue constituents and for total calcium (FC_{CAO} + FC_{CHA}), for normal breast tissue and breast lesions with and without microcalcifications. These values include both within biopsy and biopsy-to-biopsy (or patient-to-patient) variations for any one tissue class, as multiple tissue sites from each tissue class were studied on each biopsy. As might be expected, FC_{FAT} was on average the largest FC for all three tissue types (mean 81.90, 60.12 and 40.82 for normal breast tissue, breast lesions without microcalcifications and breast lesions with microcalcifications, resp.), since breast tissue contains large amounts of fat. However, in general, the mean values of the FC ranged widely both within and among the three different tissue types, reflecting differences in their morphologic and chemical composition. There was the least within class variation in FC for normal breast tissue (mean FC 0.67; range 0–1.34), which is relatively uniform in composition, consisting largely of fat. There was more within class variation in FC for breast lesions with (mean FC 12.56; range 1.55–23.27) and without microcalcifications (mean FC 7.77; range 1.53–24.35), as these categories of lesions included a range of pathology (FCC, FA and DCIS). Further each of these pathologies is itself characterized by various degrees of benign and malignant epithelial cell and stromal proliferation [37], and thus contains variable amounts of epithelial cells (nucleus and cytoplasm) and stromal collagen, the mean FC of which, for example, was greatly increased in breast lesions with (28.73) and without microcalcifications (19.73) as compared with normal breast tissue (0.34).

These results give us insight into the higher SNR in normal breast tissue than in breast lesions. This can be attributed, at least in part, to the larger spectral contribution of fat, in combination with the intrinsically high Raman scattering cross-section of fat, which results in a stronger Raman signal and so a higher SNR in normal breast tissue in comparison to the breast lesions. In addition, the presence of a higher fluorescence background in breast lesions, due to larger spectral contributions from fluorescent tissue constituents such as calcifications and collagen [38], results in a higher noise floor in these tissues. Further, breast lesions with microcalcifications exhibit on average even larger collagen contributions and smaller fat contributions and, consequently, have an even lower SNR than breast lesions without microcalcifications.

Precision of Raman Measurements

The precision, that is, the closeness of agreement between replicate Raman spectral measurements for each individual tissue site was initially assessed qualitatively by visual inspection of the replicate spectra, as shown in the typical example in Figure 1, which plots the superimposition of representative Raman spectra obtained from 5 replicate measurements from tissue sites with normal breast tissue and breast lesions (FCC) with and without microcalcifications from the same patient’s breast biopsy. There appears to be excellent precision in the Raman measurements from normal breast tissue sites, with nearly perfect superimposition of all 5 replicate Raman spectra (Figure 1a). Although the general line shape is the same, there is slightly less perfect superimposition for the 5 replicate spectra from the breast lesions without (Figure 1b) and with microcalcifications (Figure 1c), largely due to decreased SNR.

The precision of the Raman spectral measurements was also assessed quantitatively in terms of the SD and CV of the replicate model FC derived from them. It should be noted that the SD and CV of each of the 10 model FC (FC of 9 breast tissue constituents and total calcium) were first calculated for each set of 5 replicates for each individual tissue site and then the individual SD and CV for each of the 10 model FC averaged for all tissue sites of the same

tissue type. The mean replicate SD and CV for each of the 10 model FC for normal breast tissue and breast lesions with and without microcalcifications are shown in Table 2.

When assessed by SD, the precision of the Raman FC was excellent, with 99% of replicate FC values for all 10 model FC within 2SD of the mean for normal breast tissue and breast lesions with and without microcalcifications. In fact, 100% of replicate FC values were within 2SD of the mean for normal breast tissue sites and breast lesions without microcalcifications, whereas 98% were within 2SD of the mean for breast lesions with microcalcifications. However, this is not a rigorous assessment of precision given the wide variation in mean FC values within and among tissue types.

Therefore, the precision of the Raman FC was also assessed using the CV (or relative SD), which describes the standard deviation as a percentage of the mean. When assessed in terms of CV, the precision of the Raman FC was also excellent for normal breast tissue sites, with a mean CV for all 10 model FC of 0.03 (range 0–0.12). However, the CV of the model FC were somewhat higher for breast lesions without microcalcifications (mean 0.12; range 0.01–0.32), and higher still for breast lesions with microcalcifications (mean 0.22; range 0.10–0.46). There are at least two possible contributing factors. The most significant is decreased SNR in the Raman spectra from which they are obtained as discussed previously. Another possible contributing factor is that, out of necessity, three different lesion types (FCC, FA and DCIS) have been lumped together in these two broad categories (lesions with and without microcalcifications) in this work, which focuses on detection of microcalcifications and not on diagnosis of the underlying breast lesions. However, this does not appear to be a significant contributor as all lesions without microcalcifications are FCC (again mean FC 0.12). Further, although lesions with microcalcifications include normal, FCC, FA and DCIS, the mean FC for these three subgroups are similar (0.28, 0.26, 0.20 and 0.19, resp.).

Interestingly, for both breast lesions with and without microcalcifications, the CV was largest for FC_{CAO} (0.46 and 0.32, resp.); whereas the CV for FC_{CAH} was among the smallest (0.10 and 0.09, resp.). And, despite the relatively high CV for FC_{CAO}, the CV of $FC_{total\ calcium}$ (like the FC_{CAH}) was relatively small for breast lesions with and without microcalcifications (0.11 and 0.10, resp.). The CV of the FC_{FAT} and FC_{COLL} were also relatively small (≤ 0.11) for breast lesions with and without microcalcifications, which has significance for detection of microcalcifications using our Raman decision algorithms as discussed below.

Concordance of Spectral Diagnosis Among Replicates

One way to judge the significance of any imprecision in our Raman measurements from breast biopsy tissue is to assess its effect, if any, on Raman-based tissue classification/diagnosis. We previously developed two Raman decision algorithms for the detection of microcalcifications in fresh breast biopsy tissue using model FC as diagnostic parameters [21]. So we assessed the effect of the precision of our Raman measurements on concordance of the spectral diagnosis among the replicate measurements for each tissue site using each of these two algorithms.

Empirical algorithm—The first of these Raman decision algorithms for the detection of microcalcifications is an empirical algorithm based on the $FC_{total\ calcium\ (CAO+CHA)}$, using FC=11 as the decision threshold [21]. Only the first of the 5 replicate measurements was used in developing this algorithm using the initial 159 tissue site data set, resulting in a positive predictive value (PPV) of 97%, negative predictive value (NPV) of 78%, sensitivity (SE) of 77%, specificity (SP) of 97% and overall accuracy of 89% for the detection of microcalcifications.

On analysis of the replicate data, using the decision threshold of $FC_{total\ calcium}$=11, there was discordance (disagreement) in the diagnosis among the 5 replicates for only 14 of 106 (13%) tissue sites. All 28 normal breast tissue sites had complete diagnostic concordance (agreement) among the replicates for each tissue site. There was diagnostic discordance (disagreement) for 10 of 60 (17%) tissue sites with breast lesions with microcalcifications and 4 of 18 (22%) tissue sites with breast lesions without microcalcifications. The CV for $FC_{total\ calcium}$ for the 14 tissue sites with diagnostic discordance averaged 0.14, and so was not significantly different from the CV for the $FC_{total\ calcium}$ for lesions with (0.11) and without microcalcifications (0.10) as a whole. As might be expected, the 14 tissue sites with diagnostic discordance among the replicates were all line-sitters, with a mean $FC_{total\ calcium}$ of 11.12 (range 6.35–18.00), and thus were close to the decision line ($FC_{total\ calcium}$=11) when plotted in a dot plot of the empirical Raman algorithm for detection of microcalcifications, shown in red in Figure 2. Thus diagnostic discordance for these lesions is not due to greater imprecision in the Raman measurements, but rather due to FC values close to the threshold. For such line-sitters, small differences in the FC may result in misdiagnosis. However, despite being line-sitters, only 5 of the 14 tissue sites with discordance among the replicates were actually misdiagnosed during initial algorithm development using the $FC_{total\ calcium}$ of the first replicate. Four of these were false negative diagnoses for lesions with microcalcifications; one was a false positive diagnosis for a lesion without microcalcifications.

Twelve additional tissue sites were misdiagnoses during initial algorithm development using the $FC_{total\ calcium}$ of the first replicate, all false negative diagnoses for lesions with microcalcifications. In all 12 of these tissue sites, there was concordance among all 5 replicates. These 12 tissue sites had a mean value of $FC_{total\ calcium}$ of 6.88 (range 0–10.44) and, thus, are not line sitters. These misdiagnoses are most likely due to spectroscopy-histopathology registration errors.

So, in our empirical algorithm study, measurement imprecision is only an issue for line-sitters with $FC_{total\ calcium}$ close to the decision threshold. Averaging all 5 replicates corrects the misdiagnosis for 2 of the 5 line-sitter tissue sites misdiagnosed using the first replicate, but also creates misdiagnoses for 3 of the 9 line-sitter tissue sites correctly diagnosed using the first replicate. This shows that diagnosis of line-sitters was not consistently improved by averaging additional replicate measurements.

Logistic regression algorithm—The second Raman decision algorithm for the detection of microcalcifications was developed using logistic regression [21]. This is a multi-parametric algorithm, based on $FC_{total\ calcium}$, FC_{COLL} and FC_{FAT}. Again, only the first of the 5 replicate measurements was used in a leave-one-out cross-validation protocol to develop this algorithm using the initial 159 tissue site dataset, resulting in a PPV of 95%, NPV of 85%, SE of 86%, SP of 94% and overall accuracy of 90% for the detection of microcalcifications. Using this algorithm, there was discordance (disagreement) in the diagnosis among the 5 replicates for only 9 of 106 (8.5%) total tissue sites, 7 of 60 (11.7%) lesions with microcalcifications and 2 of 18 (11.1%) lesions without microcalcifications. This may be due in part to the relatively low CV ($\leq$ 0.11) for the three FC used as diagnostic parameters. Again, as might be expected, these 9 tissue sites were all line-sitters. These data points were close to the decision plane when plotted in a 3-D scatter plot constructed from an auto prediction protocol of the logistic regression algorithm for detection of microcalcifications, shown in Figure 3. It is worth noting that the leave-one-out cross-validation algorithm creates 106 different classification models and is thus not suitable for ready visualization of the respective tissue sites. Consequently the auto prediction plot is shown here to enhance our understanding of the discordant prediction points.

Again, despite being line-sitters, only 4 of the 9 tissue sites with discordance among the replicates were misdiagnosed during initial algorithm development using the $FC_{total\ calcium}$, FC_{COLL} and FC_{FAT} of the first replicate. All 4 of these were false negative diagnoses for lesions with microcalcifications. Twelve additional tissue sites were misdiagnosed during initial algorithm development using the FC of the first replicate, a majority of which (8) were false negative diagnoses for lesions with microcalcifications; the rest (4) were false positives. In all 12 of these tissue sites, there was concordance among all 5 replicates. Again, these misdiagnoses are most likely due to spectroscopy-histopathology registration errors, and measurement imprecision is only an issue for line-sitters with FC close to the decision plane. Averaging all 5 replicates corrects the misdiagnosis for only 1 of the 4 line-sitter tissue sites misdiagnosed using the first replicate. However, averaging also creates new misdiagnoses for 2 out of the remaining 5 discordant tissue sites. This again suggests that diagnosis of line-sitters is not significantly improved by averaging additional replicate measurements.

Finally, the discordance rates given above for overall diagnosis as well as diagnosis of breast lesions with and without microcalcifications (13–22% for empirical algorithm; 8.5–11% for logistic regression algorithm) can also be expressed in terms of the 95% confidence intervals for the PPV, NPV, SE and SP of each of the algorithms. These were calculated based on the FCs obtained from the replicate spectral measurements, and are: (A) empirical algorithm: PPV = [93.32%, 100%], NPV = [73.8%, 82.2%], SE = [71.94%, 82.06%] and SP = [93.36%, 100%]; and (B) logistic regression algorithm: PPV = [91.92%, 98.08%], NPV = [82.12%, 87.88%], SE = [83.02%, 88.98%] and SP = [90.12%, 97.88%].

CONCLUSIONS

To the best of our knowledge, this is the first report of the precision of Raman spectral measurements - and of the model FC derived from them - to be used in a decision algorithm for the diagnosis of breast cancer or microcalcifications. The Raman breast tissue measurements we report here are precise, judged by the superimposition of replicate Raman spectra. Statistical analysis also showed excellent precision of model FC for normal breast tissue sites (mean CV 0.03), with slightly decreased precision of model FC for breast lesions with (mean CV 0.22) and without microcalcifications (mean CV 0.12). Any imprecision in the replicate Raman spectra and model FC in breast lesions with and without microcalcifications was largely due to decreased SNR in the Raman spectra from which these model FC were obtained, which in turn was due to both decreased signal and increased noise resulting from increased background fluorescence. A detailed comparison of the precision of our Raman measurements to that of other Raman diagnostic techniques reported in the literature [32, 33] is provided in the Supporting Information Section S-3.

Repeatability, Reproducibility and Reliability

In this study, we used model FC precision to assess Raman spectral measurement repeatability, that is, the variation in measurements obtained by one person while measuring the same sample (in this case a tissue site) repeatedly. We could not test measurement reproducibility and reliability, that is, day-to-day and operator-to-operator measurement variation, as our measurements were made on fresh stereotactic breast biopsy tissue cores that had to be rapidly fixed in formalin for histopathology diagnosis following our Raman measurements. However, it should be noted that our data pre-processing protocol (wavenumber calibration, wavelength response correction, background subtraction, etc.) [39] is designed to maximize reliability by minimizing day-to-day instrument variations. The use of standard operating procedures and automation of data collection (vertical binning, acquisition time, etc.) and analysis (preprocessing, modeling, FC extraction, decision algorithm prediction, etc.) using our real-time data analysis software [39] is

designed to maximize reproducibility by minimizing operator error. The results reported here are instrument-specific to the extent that the noise levels themselves are determined by the detector noise (including dark noise, readout noise and shot noise) as well as small variations due to laser fluctuations. As such, the precision of Raman measurements will vary slightly from system to system (depending on the instrument components).

The findings of our study here provide a qualitative and quantitative assessment of the confidence of the Raman FC-based diagnosis by assignment of measurement errors, which effectively specify a probability distribution. It is worth noting that the limiting uncertainty for Raman measurements (*i.e.* in the absence of errors in the decision algorithm) stems from the overlap in the spectral components and the noise in the prediction (tissue) spectrum. The actual measurement uncertainty (and the corresponding lack of repeatability) is typically higher due to the additional uncertainty arising from an imperfect decision line. The separate contributions due to discordance among replicates (imprecision) and the misclassification of line-sitters (imperfect decision line errors) have been clearly established in the results of our *ex vivo* study.

Implications for Raman Spectroscopy Guidance of Stereotactic Breast Needle Biopsies for Microcalcifications

Here we discuss the implications of our precision measurements on the use of Raman spectroscopy as a clinical tool to provide real-time feedback to the radiologist during stereotactic needle biopsy procedures as to whether the microcalcifications seen on mammography will be retrieved. In order to implement Raman spectroscopy clinically for the detection of microcalcifications in breast biopsy tissue we must assess the influence of measurement precision on diagnostic accuracy. Overall, in our study, Raman FC precision was an issue in terms of diagnostic accuracy only for spectral diagnosis of line-sitters, that is, tissue sites with FC close to the decision line or decision plane using our previously developed FC-based Raman decision algorithms for detection of microcalcifications [21]. This is due, at least in part, to the relatively low mean CV (≤ 0.11) of the FC used as parameters in these decision algorithms, $FC_{\text{total calcium}}$, FC_{FAT} and FC_{COLL}, in all three tissue types.

It should be noted that line-sitters are a significant issue for clinical implementation of any decision algorithm or other thresholded laboratory test, regardless of measurement precision. A number of approaches could be adopted to handle diagnostic line-sitters for Raman detection of microcalcifications during stereotactic biopsy procedures. One possible approach that is generally well accepted by clinicians is to assign a confidence level to each diagnosis, for example, to render a diagnosis of lesion with microcalcifications with 95% confidence. Another possible approach is to define a region of uncertainty surrounding the decision line or decision plane. This can be done using the posterior probability of the target lesion, in this case lesions with microcalcifications, in the study population [40]. If results for a specific tissue site place it in the region of uncertainty, one could render a diagnosis of uncertain or equivocal for microcalcifications. There is precedence for the use of a decision category of equivocal, which is currently used for the fluorescence in situ hybridization (FISH) assay for Her-2 gene amplification performed on breast biopsy tissue to determine optimal therapy for invasive breast cancer [41]. Her-2 FISH assay results are expressed as the ratio of Her-2 gene copy number to gene copy number for the Cep17 gene, a marker gene for chromosome 17 on which the Her-2 gene resides. Initially a decision threshold of Her-2/Cep17 ratio=2.0 was used with two decision categories: positive for gene amplification for ratios ≥ 2.0; and negative for gene amplification for ratios < 2.0. However, br east oncologists were uncertain how to proceed therapeutically for patients with Her-2/Cep17 ratios near the decision threshold. Therefore, the guidelines for interpretation of the Her-2 FISH assay have recently been changed and now employ three decision categories:

positive for Her-2/Cep17 ratios > 2.2; negative for ratios < 1.8; and equivocal for ratios of 1.8–2.2. If the Her-2 FISH assay results are equivocal, further testing is recommended, in this case performance of an alternate immunohistochemical assay for Her-2 over-expression, in the hope of obtaining more definitive results. A similar approach could also be used for Raman diagnostic line-sitters, that is, to render an equivocal diagnosis and perform additional Raman measurements in the hope of obtaining more definitive results. To do this, one could either increase the number of frames averaged to increase overall acquisition time, decrease SNR ratio and improve model fits; or obtain additional spectra after repositioning the probe so that it is better centered over the lesion of interest. These actions require the type of real time data analysis made possible by our clinical Raman instrument.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Institute of Health National Center for Research Resources (P41-RR02594) and the National Cancer Institute (R01-CA140288).

References

1. Johnson JM, Dalton RR, Wester SM, Landercasper J, Lambert PJ. Arch Surg. 1999; 134:712–716. [PubMed: 10401820]
2. Jackman RJ, Rodriguez-Soto J. Radiology. 2006; 239:61–70. [PubMed: 16567483]
3. Radi MJ. Arch Pathol Lab Med. 1989; 113:1367–1369. [PubMed: 2589947]
4. Abramczyk H, Brozek-Pluska B, Surmacki J, Jablonska-Gajewicz J, Kordek R. Prog Biophys Mol Biol. 2011; 108:74–81. [PubMed: 22122914]
5. Keller MD, Vargis E, de Matos Granja N, Wilson RH, Mycek MA, Kelley MC, Mahadevan-Jansen A. J Biomed Opt. 2011; 16:077006. [PubMed: 21806286]
6. Krishna CM, Kurien J, Mathew S, Rao L, Maheedhar K, Kumar KK, Chowdary MV. Expert Rev Mol Diagn. 2008; 8:149–166. [PubMed: 18366302]
7. Chowdary MV, Kumar KK, Kurien J, Mathew S, Krishna CM. Biopolymers. 2006; 83:556–569. [PubMed: 16897764]
8. Chowdary MV, Kalyan Kumar K, Mathew S, Rao L, Krishna CM, Kurien J. Biopolymers. 2009; 91:539–546. [PubMed: 19226625]
9. Majumder SK, Keller MD, Boulos FI, Kelley MC, Mahadevan-Jansen A. J Biomed Opt. 2008; 13:054009. [PubMed: 19021389]
10. Bitar RA, da Martinho HS, Tierra-Criollo CJ, Zambelli Ramalho LN, Netto MM, Martin AA. J Biomed Opt. 2006; 11:054001. [PubMed: 17092150]
11. Frank CJ, Redd DC, Gansler TS, McCreery RL. Anal Chem. 1994; 66:319–326. [PubMed: 8135372]
12. Frank CJ, McCreery RL, Redd DC. Anal Chem. 1995; 67:777–783. [PubMed: 7762814]
13. Henry CM. Anal Chem. 1996; 68:718A–719A.
14. Schaeberle MD, Kalasinsky VF, Luke JL, Lewis EN, Levin IW, Treado PJ. Anal Chem. 1996; 68:1829–1833. [PubMed: 8686910]
15. Stone N, Baker R, Rogers K, Parker AW, Matousek P. Analyst. 2007; 132:899–905. [PubMed: 17710265]
16. Matousek P, Stone NJ. Biomed Opt. 2007; 12:024008.
17. Baker R, Matousek P, Ronayne KL, Parker AW, Rogers K, Stone N. Analyst. 2007; 132:48–53. [PubMed: 17180179]
18. Kerssens MM, Matousek P, Rogers K, Stone N. Analyst. 2010; 135:3156–3161. [PubMed: 20941399]

19. Stone N, Matousek P. Cancer Res. 2008; 68:4424–4430. [PubMed: 18519705]
20. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Cancer Res. 2002; 62:5375–5380. [PubMed: 12235010]
21. Saha A, Barman I, Dingari NC, McGee S, Volynskaya Z, Galindo LH, Liu W, Plecha D, Klein N, Dasari RR, Fitzmaurice M. Biomed Opt Express. 2011; 2:2792–2803. [PubMed: 22025985]
22. Shafer-Peltier KE, Haka AS, Motz JT, Fitzmaurice M, Dasari RR, Feld MS. J Cell Biochem Suppl. 2002; 87:125–137. [PubMed: 12552612]
23. Shafer-Peltier KE, Haka AS, Fitzmaurice M, Crowe J, Myles J, Dasari RR, Feld MS. J Raman Spec. 2002; 33:552–563.
24. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Shenk R, Wang N, Dasari RR, Fitzmaurice M, Feld MS. J Biomed Opt. 2009; 14:054023. [PubMed: 19895125]
25. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari RR, Crowe JP, Feld MS. Cancer Res. 2006; 66:3317–3322. [PubMed: 16540686]
26. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Proc Natl Acad Sci U S A. 2005; 102:12371–12376. [PubMed: 16116095]
27. Manoharan R, Shafer K, Perelman L, Wu J, Chen K, Deinum G, Fitzmaurice M, Myles J, Crowe J, Dasari RR, Feld MS. Photochem Photobiol. 1998; 67:15–22. [PubMed: 9477761]
28. Scepanovic OR, Bechtel KL, Haka AS, Shih WC, Koo TW, Berger AJ, Feld MS. J Biomed Opt. 2007; 12:064012. [PubMed: 18163828]
29. Arakawa M, Yamamoto J, Kagi H. Appl Spectrosc. 2007; 61:701–705. [PubMed: 17697463]
30. Mitambo MM, Zhang SL, Loppnow GR. Review of Scientific Instruments. 1998; 69:3645–3648.
31. Short SM, Cogdill RP, Anderson CA. AAPS Pharm Sci Tech. 2007; 8:E96.
32. Sowa MG, Popescu DP, Werner J, Hewko M, Ko AC, Payette J, Dong CC, Cleghorn B, Choo-Smith LP. Anal Bioanal Chem. 2007; 387:1613–1619. [PubMed: 17082878]
33. Dingari NC, Horowitz GL, Kang JW, Dasari RR, Barman I. PLoS ONE. 2012; 7:e32406. [PubMed: 22393405]
34. Vittinghoff, E.; Glidden, DV.; Shiboski, SC.; McCulloch, CE. Regression Methods in Biostatistics: linear, logistic, survival, and repeated measures models. Springer Science+Business Media, Inc; New York, NY: 2005.
35. Qi D, Berger AJ. Appl Opt. 2007; 46:1726–1734. [PubMed: 17356615]
36. Dingari NC, Barman I, Kang JW, Kong CR, Dasari RR, Feld MS. J Biomed Opt. 2011; 16:087009. [PubMed: 21895336]
37. O'Malley, FP.; Pinder, SE.; Mulligan, AM. Foundations in Diagnostic Pathology Series. 2. Elsevier Sanders; Philadelphia, PA: 2011. Breast Pathology.
38. Fitzmaurice M, Bordagaray JO, Engelmann GL, Richards-Kortum R, Kolubayev T, Feld MS, Ratliff NB, Kramer JR. Am Heart J. 1989; 118:1028–1038. [PubMed: 2816687]
39. Motz JT, Gandhi SJ, Scepanovic OR, Haka AS, Kramer JR, Dasari RR, Feld MS. J Biomed Opt. 2005; 10:031113. [PubMed: 16229638]
40. Cothren RM, Richards-Kortum R, Sivak MV Jr, Fitzmaurice M, Rava RP, Boyce GA, Doxtader M, Blackman R, Ivanc TB, Feld MS, Petras RE. Gastrointest Endosc. 1990; 36:105–111. [PubMed: 2335276]
41. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF. Arch Pathol Lab Med. 2007; 131:18–43. [PubMed: 19548375]

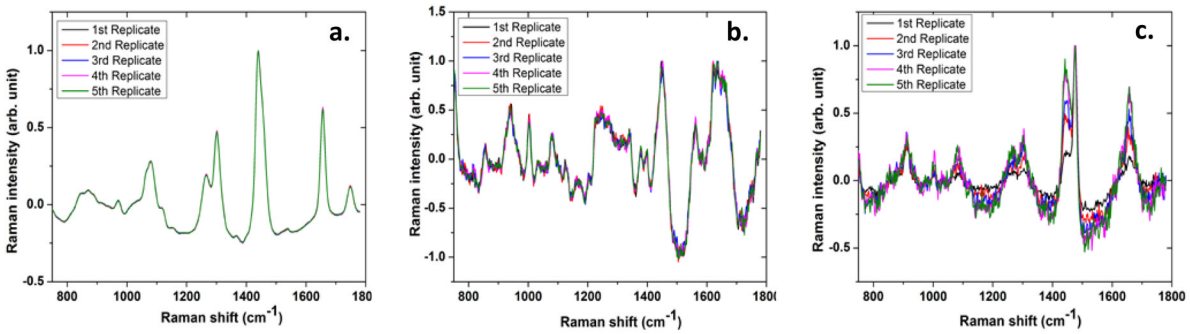

Figure 1.

Replicate Raman spectra from tissue sites with (a) normal breast tissue, (b) a lesion (FCC) without microcalcifications and (c) a lesion (FCC) with type I microcalcifications from the same patient. The 5 replicate Raman spectra from normal breast tissue in (a) show nearly perfect superimposition, indicating high precision. Although the general lineshape is the same, there is slightly less perfect superimposition for the 5 replicate spectra from the breast lesion without microcalcifications (b) and the breast lesion with microcalcifications (c), largely due to decreased SNR.

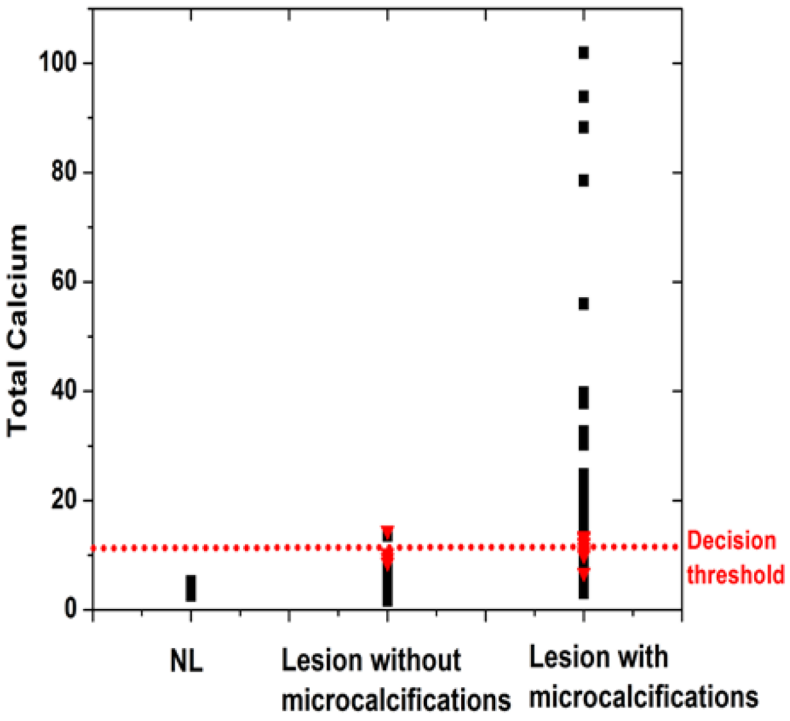

Figure 2.

Empirical Raman algorithm for detection of microcalcification based on $FC_{\text{total calcium}}$ (decision threshold=11). The 14 of 106 tissue sites with diagnostic discordance among the 5 replicate measurements (▼) are all line-sitters.

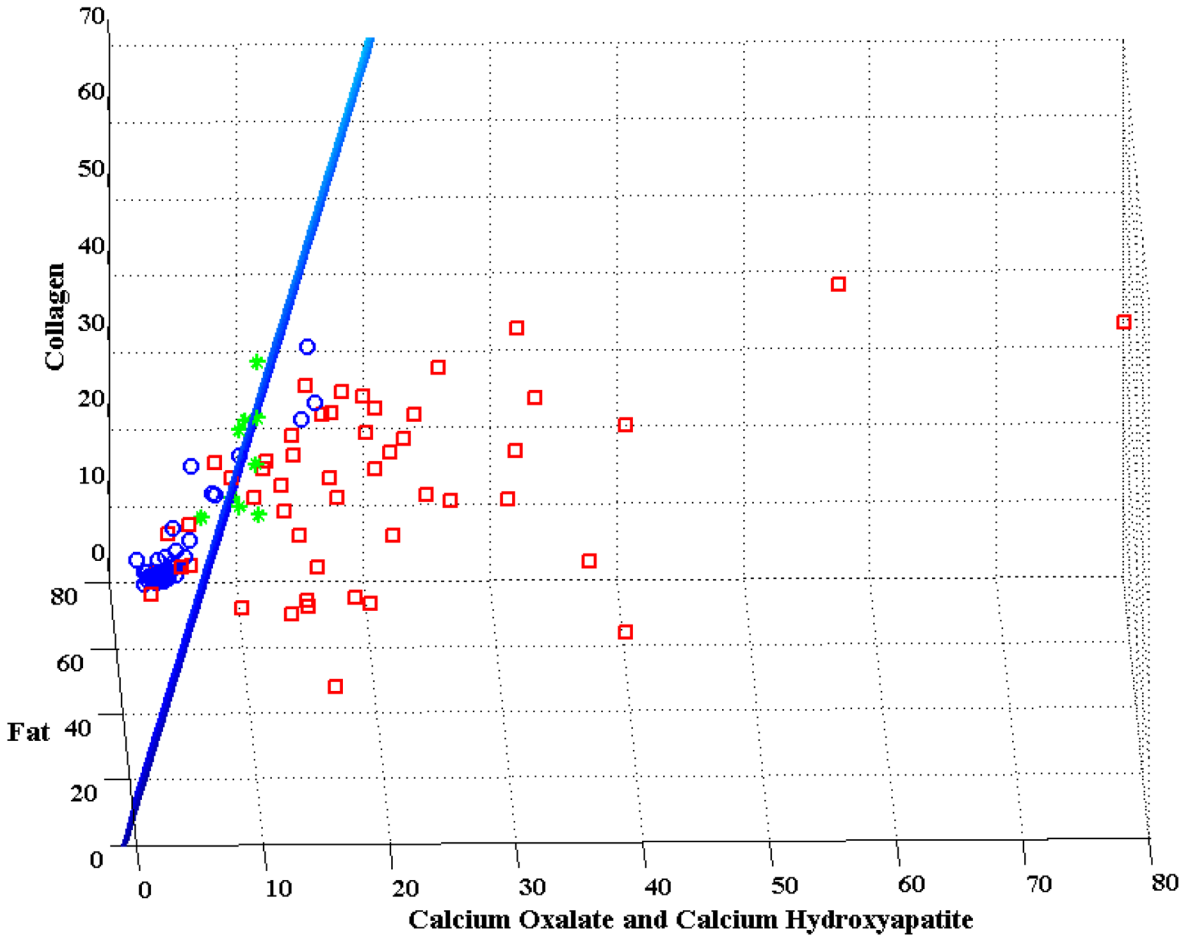

Figure 3.

Scatter plot of an auto-prediction logistic regression algorithm for distinguishing lesions with microcalcifications (□) from normal tissue samples + lesions without microcalcifications (○). Again, the 9 of 106 tissue sites with diagnostic discordance among the 5 replicate measurements (*) are all line-sitters.

Table 1

Mean FC for normal breast tissue and breast lesions with and without microcalcifications.

MEAN FC (±SD)										
TISSUE TYPE	**CAO**	**CHA**	**TOTAL CA**	**CHOL**	**β-CAR**	**FAT**	**COLL**	**ECN**	**ECC**	**OXY-HB**
NORMAL (n=28)	0	4.10 (±0.67)	4.10 (±0.67)	12.68 (±1.34)	4.20 (±1.25)	81.90 (±1.24)	0.34 (±0.50)	2.38 (±0.90)	0	0.04 (±0.20)
LESION WITHOUT MICROCALC (n=18)	0.86 (±1.53)	6.93 (±2.95)	7.79 (±3.02)	21.21 (±6.96)	4.52 (±2.57)	60.12 (±24.35)	19.73 (±18.33)	4.28 (±4.58)	0.88 (±3.73)	2.72 (±9.75)
LESION WITH MICROCALC (n=60)	4.80 (±12.46)	17.25 (±17.74)	20.05 (±21.37)	25.94 (±10.16)	5.39 (±5.39)	40.82 (±23.27)	28.73 (±16.73)	6.08 (±8.86)	5.94 (±8.06)	0.80 (±1.55)

Table 2

Replicate SD and CV for each model FC for normal breast tissue and breast lesions with and without microcalcifications.

NORMAL BREAST TISSUE (n=28)										
FC	**CAO**	**CHA**	**TOTAL CA**	**CHOL**	**β-CAR**	**FAT**	**COLL**	**ECN**	**ECC**	**OXY- HB**
REPLICATE SD (MEAN)	0	0.13	0.13	0.20	0.09	0.37	0.05	0.09	0	0
REPLICATE CV (MEAN)	0	0.04	0.04	0.02	0.02	0.005	0.12	0.04	0	0
LESIONS WITHOUT MICROCALCIFICATIONS (n=18)										
FC	**CAO**	**CHA**	**TOTAL CA**	**CHOL**	**β-CAR**	**FAT**	**COLL**	**ECN**	**ECC**	**OXY- HB**
REPLICATE SD (MEAN	0.27	0.62	0.79	1.01	0.47	1.26	0.91	0.55	0.20	0.13
REPLICATE CV (MEAN	0.32	0.09	0.10	0.04	0.19	0.05	0.06	0.25	0.01	0.13
LESIONS WITH MICROCALCIFICATIONS (n=60)										
FC	**CAO**	**CHA**	**TOTAL CA**	**CHOL**	**β-CAR**	**FAT**	**COLL**	**ECN**	**ECC**	**OXY- HB**
REPLICATE SD (MEAN)	1.09	1.33	1.97	2.57	1.04	2.12	1.33	0.83	1.20	0.24
REPLICATE CV (MEAN)	0.46	0.10	0.11	0.11	0.29	0.11	0.10	0.33	0.22	0.39