Dual-modal cancer detection based on optical pH sensing and Raman spectroscopy

Soogeun Kim
Seung Ho Lee
Sun Young Min
Kyung Min Byun
Soo Yeol Lee
Dual-modal cancer detection based on optical pH sensing and Raman spectroscopy

Soogeun Kim, Seung Ho Lee, Sun Young Min, Kyung Min Byun, and Soo Yeol Lee

Abstract. A dual-modal approach using Raman spectroscopy and optical pH sensing was investigated to discriminate between normal and cancerous tissues. Raman spectroscopy has demonstrated the potential for in vivo cancer detection. However, Raman spectroscopy has suffered from strong fluorescence background of biological samples and subtle spectral differences between normal and disease tissues. To overcome those issues, pH sensing is adopted to Raman spectroscopy as a dual-modal approach. Based on the fact that the pH level in cancerous tissues is lower than that in normal tissues due to insufficient vasculature formation, the dual-modal approach combining the chemical information of Raman spectrum and the metabolic information of pH level can improve the specificity of cancer diagnosis. From human breast tissue samples, Raman spectra and pH levels are measured using fiber-optic-based Raman and pH probes, respectively. The pH sensing is based on the dependence of pH level on optical transmission spectrum. Multivariate statistical analysis is performed to evaluate the classification capability of the dual-modal method. The analytical results show that the dual-modal method based on Raman spectroscopy and optical pH sensing can improve the performance of cancer classification. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.22.10.105002]

Keywords: Raman spectroscopy; pH sensing; dual-modal; cancer detection; diagnosis; principal component linear discriminant analysis.

1 Introduction

Over the last decades, Raman spectroscopy has been actively applied to cancer diagnosis due to its capability to detect a slight variation of chemical composition in biological tissue, caused by malignant transformation. Previous studies have shown that Raman spectroscopy can discriminate between normal and cancerous tissue in various tissues, such as cervix, skin, breast, and gastrointestinal tract. However, Raman spectroscopy has suffered from strong fluorescence background in biological applications. In addition, the subtle spectral differences between normal and disease tissues make it difficult to analyze the Raman spectra. To overcome these issues, many researchers have proposed the hardware- and/or software-based methods, such as surface-enhanced Raman scattering and data preprocessing.

While the suggestions for improving the quality of Raman spectra received the most attention in previous studies, a multimodal approach, which combines Raman spectroscopy with other methods, was also of interest to researchers. Because the malignant transformation of tissues leads to the change in metabolism and optical characteristics as well as in the chemical composition of the tissues, it is expected that we could improve the accuracy and reliability of cancer diagnosis by merging the tissue state data obtained from multimodal methods. For example, Ashok et al. reported that both sensitivity of 0.89 and specificity of 0.77 were improved to 0.94 when combining Raman spectroscopy and optical coherence tomography. Lim et al. found that a combination of three fiber-optic-based optical spectroscopy modalities of diffuse optical spectroscopy, laser-induced fluorescence spectroscopy, and Raman spectroscopy showed enhanced performance for in vivo noninvasive diagnosis of melanoma and nonmelanoma skin cancer. Kenny et al. demonstrated the potential of multimodal spectra imaging based on autofluorescence imaging and Raman microspectroscopy for realizing fast diagnosis of tumors at the surface of large surgical resections with no tissue preparation, such as sectioning and staining.

Commonly, it has been known that tumor microenvironment is hypoxic and acidic owing to the chaotic and heterogeneous microvascular structure of solid tumors. The pH level of the extracellular matrix around cancer cells has been reported to be in the range of 5.5 to 6.5. Negrini et al. revealed that the pH difference in tissue is attributed to an insufficient vasculature to remove the acidic byproducts of increased anaerobic cellular activity of the rapidly proliferating tumor cells, resulting in lower extracellular pH of tumors compared to contiguous tissues. Therefore, tissue pH level can provide useful and complementary information about cellular metabolism of the tissue in cancer diagnosis using Raman spectroscopy.

In this study, we intend to present the dual-modal approach for breast cancer diagnosis based on Raman spectroscopy and optical pH sensing. The chemical and metabolic variations in normal and cancerous breast tissue samples are measured using fiber-optic-based Raman and pH probes, respectively. For dual-modal approach, the Raman data are combined with the pH sensing data. All measured Raman data and combined data
are used to perform the classification analysis with principal component linear discriminant analysis (PC-LDA) followed by leave-one-out cross validation (LOOCV). From the results of the sensitivity and specificity obtained from the classification analysis, it is demonstrated that the dual-modal approach improves the classification capability and has the potential for more accurate breast cancer detection.

2 Materials and Methods

2.1 Sample Preparation

This study was approved by the Ethics Committee of Kyung Hee University Medical Center (KMC IRB 2016-04-103-003) and all participants signed an informed consent. Twenty breast tissue samples were collected from five participants during breast surgery. Ten samples were from the malignant tumor mass (cancer core area) and the others were from the safety margin outside of the tumor mass. Histological analysis of each sample was performed by expert pathologists and its result was used as gold standard to evaluate the performance of the proposed dual-modal cancer detection. All samples were snap frozen in liquid nitrogen after the collection and stored at −80°C to keep their metabolic content and chemical composition as in vivo state. Just before Raman and optical pH sensing experiments, each sample was thawed for about 10 min at room temperature. The size of each sample was about 5 mm × 5 mm × 1 mm.

2.2 Raman Spectroscopy

The Raman lensed fiber-optic probe (Raman probe, EmVision LLC) was used to obtain the Raman spectra from the breast tissue samples as shown in Fig. 1. The Raman probe consists of one central excitation fiber and seven Raman signal collection fibers, which have a low hydroxyl (OH) content, 300-μm core diameter and 0.22 NA. For the sample end of the Raman probe, a bandpass filter and a donut-shaped longpass filter are positioned in front of the central excitation fiber and the seven Raman signal collection fibers, respectively, to remove the excitation source signal and pass the Stokes Raman signal. The excitation source of 785 nm (IT785MM0350MF, Innovative Photonic Solutions) is delivered on the sample surface with the laser power of 200 mW through the central excitation fiber. Then, the Raman scattering signal reflected from the sample surface is collected and delivered through the seven Raman signal collection fibers and detected by a Czerny–Turner spectrograph (SR-303i-A, Andor Technology) with a low dark current deep-depletion CCD (iVac, Andor Technology).

For Raman experiments, each thawed sample was laid on a slide glass surrounded by aluminum foil. The strong and broad fluorescence signal of a slide glass caused by a 785-nm source is prevented by aluminum foil because of its infrared opacity. In addition, aluminum foil yields a very low and featureless spectral background, which enables the acquisition of high-quality Raman spectra from biological samples. Twenty Raman spectra from two different positions were measured for each breast tissue sample. All Raman spectra were acquired in the range of 600 to 1800 cm⁻¹ with the spectral resolution of 0.8 cm⁻¹ and the acquisition time of 1 s.

2.3 Optical pH Sensing

To optically measure the pH variation between normal and cancerous samples, a fiber-optic-based pH probe was made by the layer-by-layer electrostatic self-assembly technique. 3-Amino-7-dimethylamino-2-methylphenazine hydrochloride, commonly known as neutral red (NR), and poly(acrylic acid) (PAA, \(M_w \approx 1800\)) were used as the cationic chemical and anionic polyelectrolyte, respectively, for the fabrication of the nanostructured sensitive films. Ten millimolar solutions of the two chemicals was prepared using deionized water. Then, the pH of the solutions was adjusted to pH 4.2 by adding a few drops of NaOH or HCl to achieve the strong and stable electrostatic interactions between the two chemicals. All chemicals were purchased from Sigma–Aldrich Inc. and used without further purification.

A multimode plastic-clad silica-core fiber (FT200EMT, Thorlabs Inc., 200-μm core diameter, 225-μm cladding diameter, 0.39 NA) was used to fabricate the pH probe. About 1-cm part of the optical fiber cladding was chemically removed with acetone prior to the polymeric coating fabrication. The exposed part was first immersed into 1 M KOH solution for 10 min to acquire negative charge at the surface of the optical fiber core and then was rinsed in deionized water. Afterward, the exposed part was immersed into the 10 mM NR solution for 2 min to fabricate the cationic layer and rinsed in deionized water for 1 min. After the exposed part was immersed into 10 mM PAA solution for 2 min to fabricate the anionic layer and rinsed again in deionized water for 1 min. From this sequence, the bilayer of NR/PAA was fabricated at the exposed part of the optical fiber. In this study, the sequence was repeated 15 times. The transmission characteristics of the fabricated pH probe were investigated in the range of 640 to 760 nm using the standard buffer solutions of pH values 6, 7, and 8. To obtain the transmission spectra of Fig. 2(a), whenever each standard pH solution was dropped onto the surface of the exposed part of the fiber-optic-based pH probe (pH-sensitive region), the transmission spectra at pH values 6, 7, and 8 were measured by the spectrophotograph with CCD used in Raman experiment (refer to Fig. 3). The observed variation trend of the transmission spectra according to pH level is consistent with earlier study.

Figure 2(b) shows the calibration curve of the pH probe between pH values 6 and 8, which was obtained using the integrated transmission intensity values, areas under the transmission spectral curves of Fig. 2(a).
After the calibration of the pH probe, the transmission spectra of the normal and cancerous breast tissue samples were measured immediately after Raman experiments. Because Raman probe and pH probe differed in spot size and shape (Raman probe: 0.3-mm diameter, pH probe: 10 mm × 0.2 mm), the exact registration between Raman and pH measurement spots was not possible. Instead, the measurement spots of the transmission spectra overlapped with those of the Raman spots was not possible. Instead, the measurement spots of the Raman probe and pH probe differed in spot size and shape measured without the samples were simply subtracted from the raw Raman spectra of the samples. The noise-corrected Raman spectra were smoothed by the Savitzky-Golay digital filter with smoothing width of 9 and degree of 3 followed by baselining using improved modified multipolynomial fitting.13,15 To remove the spectral variations by the fluctuations of excitation source, environment, and/or sample conditions, which are not related to the actual differences in the samples, the baseline-corrected Raman spectra were normalized by the vector normalization technique,29 where each of the Raman intensities corresponding to a Raman shift is divided by the Euclidean norm. For a dual-modal approach in multivariate statistical analysis, each normalized Raman spectrum was combined with each transmission spectrum obtained from the same sample. The combined spectra (pH-Raman spectra) have the wavelength range of 640 to 760 nm for the transmission spectra and 824 to 914 nm for the Raman spectra.

The 400 normalized Raman spectra (200 from cancerous samples and 200 from normal samples) were analyzed to evaluate the classification capability of only Raman spectroscopy. Whereas the 400 pH-Raman spectra (also 200 from cancerous samples and 200 from normal samples) were analyzed for the performance evaluation of the proposed dual-modal cancer detection based on optical pH sensing and Raman spectroscopy. PC-LDA was employed for the classification analysis on only Raman spectra and the pH-Raman spectra. LOOCV was used for evaluating the performance of the classification technique. All analyses were implemented using R program version 3.3.2.

3 Results and Discussion

Figure 4(a) shows the normalized and averaged Raman spectra for the normal and cancerous breast tissue samples. Each Raman spectrum represents the average (solid and dashed lines) and standard deviation (filled areas) calculated from 200 Raman spectra. The spectral profile is consistent with earlier studies, for example, phenylalanine ring breathing (1009 cm⁻¹), phospholipids (1080 cm⁻¹), amide III in proteins (1270 cm⁻¹), CH₂ deformation in lipids (1305 and 1443 cm⁻¹), amide I in proteins (1658 cm⁻¹), and carbonyl stretch (1750 cm⁻¹).30-33 There exists spectral regions where the classification between normal and cancerous samples is possible with the naked eye. For the Raman spectra of the cancerous samples, it is observed that significant decrease in the normalized Raman intensity takes place in most of the spectral regions. In particular, the peak intensity at 1750 cm⁻¹, usually due to fat content,34 ceases to exist in the cancerous sample case. Also, the peak

2.4 Multivariate Statistical Analysis

After the measurements of Raman spectra from the breast tissue samples, the instrument and CCD noise spectra, which were
intensities at 1305 and 1443 cm$^{-1}$ assigned to vibrational modes of lipids$^{30-33}$ show much greater decrease than those at 1270 and 1658 cm$^{-1}$ assigned to vibrational modes of proteins.$^{30-33}$ That is to say that the intensity ratio of lipid to protein content is decreased for the cancerous sample case. On the other hand, the peak intensity at 1009 cm$^{-1}$, usually related to phenylalanine$^{32,33}$ which is essential amino acid of proteins, is stronger in the cancerous sample than in the normal one. The observed differences in Raman spectra between normal and cancerous samples can be thus understood to imply that normal breast tissue predominantly contains lipids, whereas cancerous breast tissue mostly contains proteins, which agrees with earlier study.$^{30,32,33}$

Figure 4(b) shows the average (solid and dashed lines) and standard deviation (filled areas) of transmission spectra ($n = 200$) obtained from the normal and cancerous breast tissue samples. For the whole transmission spectral region, it is observed that the average transmission intensities of normal samples are stronger than those of cancerous samples. These observations can be explained through the absorbance change of NR/PAA bilayer at varying pH level. Goicoechea et al.$^{28}$ reported that the absorbance of NR/PAA bilayer decreases as the pH level is raised within the wavelength range from 600 to 800 nm. This, in other words, means that the transmission intensity by NR/PAA bilayer increases with increasing pH level. Therefore, from Fig. 4(b), it is confidently expected that the pH level of normal breast tissue samples is higher than that of cancerous ones. The integrated transmission intensity values of normal and cancerous samples were $4.335 \times 10^7 \pm 0.176 \times 10^7$ and $4.133 \times 10^7 \pm 0.12 \times 10^7$, respectively. Using the equation of pH calibration curve of Fig. 2(b), the pH levels of normal and cancerous samples were estimated to be about $7.37 \pm 0.9$ and $6.34 \pm 0.62$, respectively, which corresponds to the well-known fact that the pH level of cancerous tissues is lower than that of normal tissues and lies within the pH range from 5.5 to 6.5.$^{20-22}$

Principal component analysis (PCA) was used to reduce the high dimensionality in the spectral data of the normal and cancerous samples. Figure 5 shows the cluster plot of PCA scores on the principal components (PCs) of normal and cancerous samples in Raman-only and pH-Raman spectral analysis. The first, second, and third PCs in Raman-only spectral analysis accounted for 82.7%, 2.6%, and 2.1% of the total variance, respectively, whereas those in pH-Raman spectral analysis were 73.5%, 7.4%, and 4.7%. In Fig. 5, it seems that PCA can successfully separate the cluster groups between normal and cancerous samples regardless of the use of pH data. To more accurately assess the classification capability between the Raman-only and pH-Raman detection method, PC-LDA followed by LOOCV was used. The number of PCs is determined using two selection criteria: scree plot test and cumulative percentage variance.$^{34}$ Figure 6 shows the scree plot of the eigenvalues (variances) along the PC number in Raman-only and pH-Raman spectral analysis. The scree plot test searches for a knee point in the plot curve where the optimal number of PCs reached.$^{34}$ From the Raman-only curve in Fig. 6, the third PC can be selected as a knee point for Raman-only
spectral analysis. On the contrary, for pH-Raman spectral analysis, it is difficult to determine the knee point due to the smooth curve. In cumulative percentage variance, the optimal number of PCs is determined using a user-defined threshold value of cumulative percentage of the variance (usually 90% to 99% of the variance). In this study, 90% was used as the threshold value. When the number of PCs was 5 (PC 1 to 5), the cumulative percentage of the variance was 90.5% and 90.4% for Raman-only and pH-Raman spectral analysis, respectively. Therefore, the PC 1 to 5 was used to perform classification analysis using PC-LDA with LOOCV.

The results of multivariate statistical analysis using the PC 1 to 5 are represented in Table 1. The confusion matrix shows the tabular data for predicted class versus actual class of normal and cancerous samples in Raman-only, pH-only, and pH-Raman spectral analysis. The results of sensitivity and specificity are equal to 1. On the other hand, the specificity improved from 0.91 to 0.98 for pH-Raman case. Enhanced specificity quantitatively presents the potential of dual-modal cancer detection using fiber-optic-based Raman and pH probes for accurate diagnosis.

![Fig. 6 Scree plot of the eigenvalues along the principal component number analyzed using Raman-only and pH-Raman spectra of the normal and cancerous breast tissue samples. The scree plot test is used to determine the optimal number of PCs applied to the multivariate statistical analysis.](image)

### Table 1 Confusion matrix of Raman-only, pH-only, and pH-Raman spectral analysis.

<table>
<thead>
<tr>
<th></th>
<th>Raman-only</th>
<th>pH-only</th>
<th>pH-Raman</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(actual)</td>
<td>200 (TP)</td>
<td>17 (FP)</td>
<td>162 (TP)</td>
</tr>
<tr>
<td>(predicted)</td>
<td>162 (TP)</td>
<td>34 (FP)</td>
<td>200 (TP)</td>
</tr>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(actual)</td>
<td>0 (FN)</td>
<td>183 (TN)</td>
<td>166 (TN)</td>
</tr>
<tr>
<td>(predicted)</td>
<td>0 (FN)</td>
<td>183 (TN)</td>
<td>166 (TN)</td>
</tr>
</tbody>
</table>

Note: TP, true positive; TN, true negative; FP, false positive; and FN, false negative.

## 4 Conclusion

From the experimental study of measuring the Raman spectra and pH levels of normal and cancerous breast tissue samples, it was found that both Raman spectra with chemical information and transmission spectra with pH information are useful in classifying normal and cancerous samples. When the dual-modal method based on Raman spectroscopy and optical pH sensing and multivariate statistical analysis work together, it is demonstrated that classification between normal and cancerous breast tissue samples was significantly improved. More specifically, multivariate statistical analysis using PC-LDA followed by LOOCV showed an enhanced specificity for breast cancer diagnosis. The suggested approach supports the potential practical applications of pH-Raman-based detection method for breast cancer diagnosis.

## Disclosures

All authors declare no conflicts of interest or financial relationships to disclose.

## Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (2015M2A2A7A03043177 and 2017R1A2B4012428).

## References

Kim et al.: Dual-modal cancer detection based on optical pH sensing and Raman spectroscopy


Soogeon Kim received his MS and PhD degrees in mechatronics from Gwangju Institute of Science and Technology in 2007 and 2015, respectively. He is a research professor at Kyung Hee University, Republic of Korea. His research interests focus on laser–tissue interaction phenomena and their applications, and ongoing research includes Raman spectroscopy for cancer detection, fiber-optic-based Raman spectroscopy, and spectrum data analysis by neural networks.

Seung Ho Lee received his MS degree in biomedical engineering from the Department of Medical Engineering, Kyung Hee University, in 2013. Since 2013, he has studied biomechatronics and biosensors with biomedical engineering on a doctorate path in the Department of Medical Engineering, College of Medicine at Kyung Hee University.

Sun Young Min received her BS degree in medicine from Chungnam National University in 2003 and her MS and PhD degrees in medicine from Kyung Hee University in 2008 and 2011, respectively. She is an assistant professor at Kyung Hee University. Her current research interests include breast cancer, cancer pathology, and convergence technology research. She is a member of the Korean Breast Cancer Society.

Kyung Min Byun received his BS and MS degrees from the School of Electrical Engineering, Seoul National University, Seoul, Republic of Korea, and his PhD from Seoul National University in 2007. From July 2007 to February 2008, he worked as a visiting scientist in the Department of Biomedical Engineering, Cornell University, Ithaca, New York. Currently, he is working as an associate professor in the Department of Biomedical Engineering, Kyung Hee University, Yongin, Republic of Korea. His main research activities are theoretical and experimental studies on high-sensitivity plasmonic biosensors with metallic nanostructures, Raman spectroscopy, and minimally invasive neuromodulation technique based on optical and ultrasound stimulation.

Soo Yeol Lee received his MS and PhD degrees in electronic engineering from Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea, in 1985 and 1989, respectively. He was in the Department of Biomedical Engineering, Konkuk University, Republic of Korea, from 1992 to 1999. In 1999, he joined the Department of Biomedical Engineering, Kyung Hee University, where he is currently the director of the Functional and Metabolic Imaging Research Center. His current research interests include MRI, CT, elastography, and medical image/signal processing.