Evaluation of the diagnostic potential of ex vivo Raman spectroscopy in gastric cancers: fingerprint versus high wavenumber

Xueqian Zhou
Jianhua Dai
Yao Chen
Guangjie Duan
Yulong Liu
Hua Zhang
Hongbo Wu
Guiyong Peng

Evaluation of the diagnostic potential of ex vivo Raman spectroscopy in gastric cancers: fingerprint versus high wavenumber

Xueqian Zhou, Jianhua Dai, Yao Chen, Guangjie Duan, Yulong Liu, Hua Zhang, Hongbo Wu, and Guiyong Peng

Abstract. The aim of this study was to apply Raman spectroscopy in the high wavenumber (HW) region (2800 to 3000 cm\(^{-1}\)) for ex vivo detection of gastric cancer and compare its diagnostic potential with that of the fingerprint (FP) region (800 to 1800 cm\(^{-1}\)). Raman spectra were collected in the FP and HW regions to differentiate between normal mucosa (\(n=38\)) and gastric cancer (\(n=37\)). The distinctive Raman spectral differences between normal and cancer tissues are observed at 853, 879, 1157, 1319, 1338, 1448, and 2932 cm\(^{-1}\) and are primarily related to proteins, lipids, nucleic acids, collagen, and carotenoids in the tissue. In FP and HW Raman spectroscopy for diagnosis of gastric cancer, multivariate diagnostic algorithms based on partial-least-squares discriminant analysis, together with leave-one-sample-out cross validation, yielded diagnostic sensitivities of 94.59% and 81.08%, and specificities of 86.84% and 71.05%, respectively. Receiver operating characteristic analysis further confirmed that the FP region model performance is superior to that of the HW region model. Better differentiation between normal and gastric cancer tissues can be achieved using FP Raman spectroscopy and PLS-DA techniques, but the complementary nature of the FP and HW regions make both of them useful in diagnosis of gastric cancer.

Keywords: Raman spectroscopy; high wavenumber; fingerprint; gastric cancer.

1 Introduction

Decreasing incidence and mortality rates have been reported for gastric cancer in most developed countries and also in selected developing countries over the last few decades. Still, 723,100 people died of gastric cancer, 951,600 new cases occurred worldwide in 2012\(^{[9]}\) and >40% of these patients were Chinese.\(^{[10]}\) Early detection and effective treatment of gastric cancer have a great impact on the 5-year survival rates of patients,\(^{[11]}\) but unfortunately, even in Japan (which is a highly developed Asian country), difficulties occur in early detection of gastric cancer because of a shortage of skilled endoscopists and limited availability of endoscopic devices.\(^{[12]}\) Not to mention the problems and possibilities of nationwide screening for gastric cancer in China, which has the largest population in the world and great imbalances in regional development. Histological examination of gastric biopsy has been widely used since the popularization of endoscopy,\(^{[13]}\) which remains the gold standard for gastric cancer diagnosis. However, gastric biopsy under direct vision suffers from several disadvantages, such as lack of nondestructive character, extension of endoscopic examination time, and a delay of several days for the diagnostic result provided to patients. Consequently, additional efforts should be focused on the development of endoscopic diagnostic techniques that are real-time, nondestructive, and objective.

First introduced by Raman in 1928, Raman spectroscopy has shown great potential in diagnosis.\(^{[14]}\) This technique detects an emission spectrum generated by high-frequency inelastic scattering of light resulting from the response of a molecule to incident rays.\(^{[15]}\) Each molecular motion creates its own individual bands in the Raman spectrum.\(^{[16]}\) Hence, Raman spectroscopy reveals biomolecular changes in the human body caused by malignancy and presents an opportunity for use as a promising tool for doctors in cancer diagnosis. In recent years, Raman spectroscopy has been used to diagnose different types of cancers and precancers in vivo and ex vivo, including glioma, nasopharyngeal carcinoma, Barrett’s esophagus, lung cancer, breast cancer, and melanoma.\(^{[17]}\) The sensitivity and specificity for diagnosis of malignancies reported in these research studies are satisfactory and encouraging.\(^{[18]}\)

The Raman probe is sufficiently slender to freely pass through the instrument channel of an endoscope. Gastric biopsy is a routine process for patients suspected of suffering from gastric cancer, and this technique makes it easier to obtain samples compared with parenchymal viscera (e.g., brain, liver). Many doctors and optical scientists attempted to improve the diagnostic effect of Raman spectroscopy for gastrointestinal tumors both in vivo and ex vivo.\(^{[19]}\) However, most of these studies focus on the fingerprint (FP) Raman spectroscopy.
range of 800 to 1800 cm\(^{-1}\), which presents biochemical information on the backbone structures of DNA, RNA, proteins, lipids, collagen, and other biological macromolecules. In 2014, we also used the FP region of Raman spectroscopy to explore the biochemical characteristics of molecules associated with the malignant transformation of gastric mucosa.\(^\text{2c}\) In our previous study, the great value of the high wavenumber (HW) region (2400 to 3800 cm\(^{-1}\)) in diagnosis was not yet evaluated and discussed.\(^\text{21}\) Compared with the main advantages of the HW region, in which the intense background signal of the silica fiber-optic probe is absent and more intense tissue Raman signals are generated, it appears that the FP region provides much more information on the molecular composition and structure of tissue.\(^\text{20}\) However, the background-signal-free design of the Raman probe for FP Raman spectroscopy is much more complicated and expensive,\(^\text{21}\) which could further affect development and medical application of in vivo diagnostic tools based on Raman spectroscopy.\(^\text{2}\) Therefore, it is worthwhile to comprehensively investigate the diagnostic effect of both the FP and HW regions.

In this research, we compared the diagnostic potential of FP (800 to 1800 cm\(^{-1}\)) Raman spectroscopy with that of HW (2800 to 3000 cm\(^{-1}\)) Raman spectroscopy in gastric cancers. The biochemical changes of molecules found in both the FP region and the HW region were thoroughly reviewed as well. Multivariate statistical techniques, including partial least-squares-discriminant analysis (PLS-DA), were used to design diagnostic algorithms for discrimination between normal mucosa and gastric cancer.

### 2 Materials and Methods

#### 2.1 Raman Instrumentation and Spectral Measurements

The Raman spectra of gastric biopsy samples were recorded using a micro-Raman spectroscopy system (inVia Reflex, Renishaw, UK). A 785-nm laser (maximum output of 300 mW, Renishaw, UK) was used as the excitation source, with a power of 10% and an integration time of 10 s. The instrument was equipped with a Leica 50x objective. The laser power on the surface of the samples was ~20 mW, and the resolution of the spectroscopy technique is 2 cm\(^{-1}\). To ensure that the instrument was operating properly, silicon was used to check the spectrum prior to recording. Samples were removed from the storage vial and thawed at room temperature. Each sample was placed on an aluminum holder and measured under the same conditions. Two to seven observation fields were randomly selected, according to the size of each tissue sample. The average was used to represent the Raman spectrum of the sample.

#### 2.2 Patients and Biopsy Sample Collections

Biopsy samples of gastric cancer were collected from patients suspected of suffering from gastric cancer during routine upper gastrointestinal endoscopic exams at Southwest Hospital, first affiliated hospital of Third Military Medical University. Biopsy samples of normal tissue were collected from healthy volunteers in the same hospital who were free from any other diseases or treatments. Informed consent was obtained from the patients and healthy volunteers prior to sample collection. This study was approved by Southwest Hospital Ethics Committee, in agreement with the Helsinki declaration.

### Table 1: Patient characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Normal (n = 38)</th>
<th>Gastric cancer (n = 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>43.4 ± 11.8</td>
<td>63.2 ± 9.3</td>
</tr>
<tr>
<td>Gender (F/M)</td>
<td>21/17</td>
<td>10/27</td>
</tr>
<tr>
<td>Biopsy location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper third</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Middle third</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>Lower third</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Pathological diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffuse type</td>
<td>—</td>
<td>9</td>
</tr>
<tr>
<td>Intestinal type</td>
<td>—</td>
<td>28</td>
</tr>
</tbody>
</table>

methods were carried out in accordance with the approved guidelines. No patients received any treatments such as chemotherapy or radiotherapy prior to sample collection. Tissue specimens were snap frozen in liquid nitrogen until analysis. After the Raman measurements were completed, the tissue samples were fixed with 10% formalin and pathologically confirmed. In all, 38 normal and 37 gastric cancer biopsy samples (each from different patients and healthy volunteers) were included in this research. The characteristics of the patients and healthy volunteers are listed in Table 1. The histological type was categorized as diffuse type or intestinal type according to the Lauren classification.

#### 2.3 Data Processing

The Raman spectrum in the wavenumber range of 400 to 3000 cm\(^{-1}\) was acquired from all tissue samples to analyze and compare the diagnostic effect of FP Raman spectroscopy with that of HW Raman spectroscopy. In this research, our analyses were limited to the FP region (800 to 1800 cm\(^{-1}\)) and the HW region (2800 to 3000 cm\(^{-1}\)) because these two regions provide the greatest biochemical information from human tissue.\(^\text{2}\) The acquired raw Raman spectra consisted of tissue autofluorescence background and noise, therefore, a second-order Savitzky–Golay filter (window width of 7 pixels) was used to reduce the noise. Subsequently, the autofluorescence background in the noise-smoothed spectrum was subtracted using an algorithm known as adaptive iteratively reweighted penalized least squares (airPLS) that does not require any user intervention or prior information.\(^\text{2}\) To compare the spectral shapes and relative Raman band intensities between the different samples, all of the acquired Raman spectra were normalized to the integrated area under the curve (FP or HW). All data were processed in the MATLAB environment (The MathWorks, Inc., Natick, Massachusetts).

#### 2.4 Statistical Analysis

Probabilistic PLS-DA, which was reported to be more effective than principal component analysis (PCA) in cell and tissue classification, was applied to evaluate the difference between normal gastric mucosa and gastric cancer.\(^\text{2}\) The leave-one-sample-out cross-validation method was used to confirm...
the performance of the PLS-DA diagnostic algorithm. The number of retained latent variables (LVs) was determined based on the error rate of the cross-validation curves. The receiver operating characteristic (ROC) curves were used to further compare the performance of the PLA-DA models using the two spectral data sets (FP, HW) for ex vivo diagnosis of gastric cancer.

3 Results

Comparison of the mean intensities of FP and HW Raman spectra with their standard deviations (SD) is shown in Fig. 1. The prominent tissue Raman peaks in the FP region are observed at 829, 853, 879, 937, 1003, 1033, 1086, 1127, 1157, 1173, 1207, 1319, 1338, 1448, 1526, and 1658 cm$^{-1}$. Tentative vibrational mode assignments for the human gastric tissue Raman peaks in the FP region are summarized in Table 1 of our previous work. Moreover, intense Raman peaks are also found in the HW region, e.g., 2849 cm$^{-1}$ (symmetric CH$_2$ stretching of lipids), 2875 cm$^{-1}$ (CH$_2$ symmetric stretching of lipids and proteins), and 2932 cm$^{-1}$ (CH$_3$ asymmetric stretching of lipids and proteins).

Independent-sample $t$-tests showed that Raman peak intensities are notably higher for gastric cancer at 1157 cm$^{-1}$ (polyene chain of carotenoids), 1173 cm$^{-1}$ [$\delta$C–H in-plane bending of phenylalanine, tyrosine (protein)], 1319 and 1338 cm$^{-1}$ (oCH$_2$CH$_3$ and $\delta$CH$_2$CH$_3$ of proteins and nucleic acids), and 1526 cm$^{-1}$ ($\nu$C–C of carotenoids), but notably lower differences are significant at 829 cm$^{-1}$ ($\nu$as PO$_2$ of nucleic acid), 853 cm$^{-1}$ [$\nu$C–C of proline (collagen)], 879 cm$^{-1}$ [$\nu$C–C hydroxyproline (collagen)], 1003 cm$^{-1}$ (ring breathing of phenylalanine), 1033 cm$^{-1}$ ($\delta$C–H plane bending of aromatic compounds), 1448 cm$^{-1}$ ($\delta$CH$_2$ of proteins and lipids), 1658 cm$^{-1}$ [amino compounds I (a-helix)], and 2932 cm$^{-1}$ (CH$_3$ stretching of proteins). To better reflect the differences in the main Raman peaks between normal tissue and gastric cancer, a comparison of the mean intensities and SDs is shown in Fig. 2.

The PLS-DA with leave-one-sample-out cross validation was applied to generate diagnostic algorithms for differentiation between normal tissue and gastric cancer. As shown in Fig. 3, six LVs were found to be the optimal number of retained components for PLS-DA modeling of both the FP and HW regions.
components for PLS-DA modeling of both the FP region and HW region as defined by the minimum error rate of cross-validation. To demonstrate the use of LVs for diagnosis and the intrinsic clustering of normal mucosa and gastric cancer, two three-dimensional (3-D) scatterplots are presented in Fig. 4 using the three LVs (LV1, LV2, and LV3) of the FP and HW regions, respectively. The LV points are divided into two distinct groups, suggesting that both FP and HW Raman spectroscopy have potential for use as a clinical diagnosis tool for gastric cancer. The developed PLS-DA algorithm of the FP region achieves diagnostic sensitivities of 94.59%, specificities of 86.84%, and accuracies of 90.67%. The HW modeling achieves diagnostic sensitivities of 83.78%, specificities of 71.05%, and accuracies of 77.33%. The ROC curves (Fig. 5) are also generated for the two spectral datasets (FP, HW) to further evaluate the diagnostic performance. The integration areas under the ROC curves for FP modeling are 0.9495, and for HW modeling, the number is 0.8254. The diagnostic performance comparisons (i.e., sensitivities, specificities, accuracies, and AUCs) of the FP and HW PLS-DA modeling are summarized in Table 2.

4 Discussion

Raman spectroscopy facilitates the collection of specific molecular information for tissue diagnosis and characterization in a nondestructive manner. This distinctive vibrational spectroscopic technique allows researchers to study the tissue in its natural condition without the need for tissue preparation. Currently, Raman studies of gastric cancer have evaluated Raman spectra mostly in the FP region. Only a few articles have reported Raman measurements from both the FP and HW regions but multivariate statistical analysis was not performed in these studies to evaluate the diagnostic potential of the FP and HW regions. In this work, we compared the diagnostic potential of FP Raman spectroscopy in gastric cancers with that of HW Raman spectroscopy and investigated the biochemical information contained in both spectral regions for the purpose of tissue characterization.

When normal tissues experience the process of malignant disease initiation, its structural conformation and relative

Table 2 Comparison of diagnostic performance of different Raman spectral regions (FP, HW) for differentiation of gastric cancer from normal mucosa.

<table>
<thead>
<tr>
<th>Raman spectroscopic modalities</th>
<th>Retained LVs</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy (%)</th>
<th>ROC (AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP</td>
<td>6</td>
<td>94.59 (35/37)</td>
<td>86.84 (33/38)</td>
<td>90.67 (68/75)</td>
<td>0.9495</td>
</tr>
<tr>
<td>HW</td>
<td>6</td>
<td>83.78 (31/37)</td>
<td>71.05 (27/38)</td>
<td>77.33 (58/75)</td>
<td>0.8254</td>
</tr>
</tbody>
</table>

Abbreviations: FP, fingerprint; HW, high wavenumber; LVs, latent variables; ROC, receiver operating characteristic; AUC, area under curve.
content of macromolecules are altered. The main advantage of using the FP region is that it provides highly specific information on proteins, lipids, and nucleic acids associated with malignancy. For instance, in a previous study, we reported that the peak at 1526 cm\(^{-1}\) was significantly stronger in gastric cancer tissue than normal tissue \((p < 0.05)\), indicating that the carotenoid content in cancer is increased. In this work, another peak at 1157 cm\(^{-1}\) (polyene chain of carotenoids) was also found to be significantly stronger in gastric cancer tissue \((p < 0.05)\). These observations provide further evidence for our assumptions that cancer cells probably evolve an enhanced ability to resist damage because of the antioxidant function of carotenoid. In addition, the peak at 853 cm\(^{-1}\) \([\text{C} – \text{C} \text{ of proline (collagen)}]\) was observed to be lower for gastric cancer tissue. A possible reason for this result is that cancer cells synthesize and secrete matrix metalloproteinase that degrades matrix proteins such as collagen to assist cancer metastasis. Huang et al. also observed a decrease of collagen in gastric cancer in vivo using a novel image-guided Raman endoscopy technique. In addition, the protein and nucleic acid Raman signals at 1319 and 1338 cm\(^{-1}\), which most likely arise from the chromosomes in the nucleus, have been found to be significantly higher for gastric cancer tissues compared with normal tissue. This observation might be explained by an increase in nuclear chromatin and excessive DNA replication during malignant transformation. Therefore, the distinct differences in the FP region between normal tissues and gastric cancer reinforce the hypothesis that the FP region can be used to interpret the biochemical transformation of molecules associated with gastric mucosa malignancy.

However, the HW region (2800 to 3000 cm\(^{-1}\)) appears to contain less biochemical information than the FP region. Raman peaks at 2849, 2875, and 2932 cm\(^{-1}\) are observed in the HW region. The protein signal at 2932 cm\(^{-1}\) was significantly stronger in normal tissue than in the gastric cancer tissue \((p < 0.05)\). The other two peaks showed no significant difference. However, this result does not mean that the HW region is less important for diagnosis. The HW region contains information related to CH\(_3\) stretching of proteins, which are complementary to the properties of the FP region. In addition, the biochemical information provided by the HW region is sufficient for cancer diagnosis. Nijssen et al. reported that a linear discriminant analysis (LDA)-based tissue classification model using the HW region (2800 to 3125 cm\(^{-1}\)) discriminated between basal cell carcinoma (BCC) and normal skin with high accuracy.

We further used PLS-DA with leave-one-sample-out cross validation to generate diagnostic algorithms for gastric cancer using the FP region and the HW region. The PLS-DA modeling of the FP region, which achieved diagnostic accuracies of 90.67\% (68/75) [sensitivities of 94.59\% (35/37) and specificities of 86.84\% (33/38)] performed better than the HW region modeling. The HW modeling misclassified 17 samples out of the total 75 samples, but the FP modeling reduced the number of misclassified samples to 7. Similar results were also observed in cervical precancers that were investigated using simultaneous FP (800 to 1800 cm\(^{-1}\)) and HW (3100 to 3700 cm\(^{-1}\)) confocal Raman spectroscopy. The ROC curves (Fig. 5) further validate the better performance and robustness of the FP region modeling. Fortunately, compared with the accuracy of gastric biopsy under direct vision, which was reported to vary from 50\% to 96\%, the performances of both FP modeling and the HW modeling are satisfactory for diagnosis of gastric cancer. Furthermore, we also performed the joint analysis of these two regions, but the diagnostic power is the same with the FP modeling. The FP + HW modeling also achieves diagnostic sensitivities of 94.59\%, specificities of 86.84\%, and accuracies of 90.67\%.

Although our work showed that FP region modeling provided diagnostic advantages over HW region modeling, the development of both FP region and HW region diagnostic technologies should be encouraged because the HW region contains information related to both lipids, proteins, and water bands that does not show up in the FP region. Another reason is that the great difference in composition of human tissue (e.g., brain and bone) makes it worthwhile to find a suitable region of Raman spectroscopy for use in building diagnostic algorithms. In addition, it has been reported that diagnostic algorithms developed ex vivo might not be directly used for in vivo applications in the colon. The performance comparison of the FP and the HW diagnostic algorithms developed for gastric cancer in vivo is not complete, therefore, further investigations are needed to determine whether the performance of the FP region modeling is still better than the HW region modeling in vivo. Another limitation of this study was the constraint on experimental conditions which means that the sample size of our study might not be sufficiently large to provide better PLS-DA modeling of both the FP region and HW region. We plan to enlarge the sample size to improve the diagnostic algorithms in our future studies. Currently, we are working on prototype spectrometers with a Raman probe that is compatible with the instrument channel of the endoscope for in vivo applications. The ultimate goal of our work is to develop a Raman spectroscopy diagnosis system that can be used as a diagnostic tool during screening for gastric cancer and will also provide an accurate gastrointestinal tumor margin assessment during endoscopic submucosal dissection. It is difficult to predict whether Raman spectroscopy can replace histopathology, but we do believe that this approach will provide good clinical tools for endoscopists in the future.

5 Conclusions

We compared the diagnostic potential of FP Raman spectroscopy in gastric cancers with that of HW Raman spectroscopy ex vivo and uncovered the biochemical information contained in both spectral regions for the purpose of tissue characterization. Better differentiation between normal and gastric cancer tissues can be achieved using FP Raman spectroscopy and PLS-DA techniques. However, the HW region contains complementary spectral information that might have a positive effect for gastric cancer diagnosis. We expect that both FP Raman spectroscopy and HW Raman spectroscopy will become useful clinical diagnostic tools for the early detection of gastric cancer in the future.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 81272580) and Chongqing Science & Technology Commission (Grant No. cstc2015shmszx10017). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
References


Xueqian Zhou received his BS degree from Hubei Medical College in 2008 and his MD degree from the Third Military Medical University in 2016. He is a PhD student in the Department of Gastroenterology, Southwest Hospital of Third Military Medical University, Chongqing, China. His current research interests include diagnosis of digestive diseases and clinical application of Raman spectroscopy.

Jianhua Dai received her MD degree from the Third Military Medical University, China, in 2009. Currently, she is an attending physician at Southwest Hospital (The First Affiliated Hospital of Third Military Medical University), where she is pursuing her doctor’s degree. Her working and research field is diagnosis and treatment of digestive diseases.

Yao Chen received her BS degree from Chongqing Medical University in 2011 and her MD degree from the Third Military Medical University in 2014. Then she became a resident at Southwest Hospital. She majored in diagnosis and treatment of gastrointestinal diseases through gastrointestinal endoscopy. Her current research mainly focused on Raman spectroscopy of gastrointestinal tumors.

Guangjie Duan received his PhD in medicine from the Third Military Medical University in 2010. He is a clinical pathologist working in the Institute of Pathology of Southwest Hospital. He is the author of more than 20 publications in pathology of tumor and inflammation. His current research interests include the pathological diagnosis and molecular study of disease of digestive and urologic system.

Yulong Liu is a professor and director of the Research Center of Intelligent Equipment and Instrumentation at Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences. He is the author of more than 180 publications. He was president of the 7th Committee on Light Scattering Branch of Chinese Physical Society. Currently, his main research interests are Raman spectroscopy and Brillouin scattering technology as well as instrumentation development.
Hua Zhang received her MD degree in chemistry from Tianjin University in 2012. She is a research assistant of the Research Center of Intelligent Equipment and Instrumentation at Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences. Her research concerns nanomaterials and their applications in surface enhanced Raman scattering (SERS). She has published 14 papers in international journals.

Hongbo Wu received his BS and MD degrees in clinical medicine from the Third Military Medical University. His research aims to prevent esophageal stricture after endoscopic submucosal dissection and clinical application of Raman spectroscopy.

Guiyong Peng received his BS and MD degrees from the Third Military Medical University. He is a professor and director of the Digestive Endoscopy Center at Southwest Hospital of Third Military Medical University, Chongqing, China. He is the author of more than 130 publications. Currently, he is on the Standing Committee of Chinese Digestive Endoscopy Society. His professional expertise lies in the diagnosis and treatment of gastrointestinal cancer through endoscopy.