Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Discrimination of premalignant lesions and cancer tissues from normal gastric tissues using Raman spectroscopy

Shuwen Luo
Changshui Chen
Hua Mao
Shaoqin Jin

SPIE

Discrimination of premalignant lesions and cancer tissues from normal gastric tissues using Raman spectroscopy

Shuwen Luo,[a*] Changshui Chen,[a*] Hua Mao,[b] and Shaoqin Jin[b]
[a]South China Normal University, Key Laboratory of Laser Life Science of Ministry of Education, Photon and Nano Research Centre for Biosciences, Guangzhou, Guangdong 510631, China
[b]Southern Medical University, Zhujiang Hospital, Department of Gastroenterology, Guangzhou, 510282 Guangdong, China

Abstract. The feasibility of early detection of gastric cancer using near-infrared (NIR) Raman spectroscopy (RS) by distinguishing premalignant lesions (adenomatous polyp, $n = 27$) and cancer tissues (adenocarcinoma, $n = 33$) from normal gastric tissues ($n = 45$) is evaluated. Significant differences in Raman spectra are observed among the normal, adenomatous polyp, and adenocarcinoma gastric tissues at 936, 1003, 1032, 1174, 1208, 1323, 1335, 1450, and 1655 cm^{-1}. Diverse statistical methods are employed to develop effective diagnostic algorithms for classifying the Raman spectra of different types of *ex vivo* gastric tissues, including principal component analysis (PCA), linear discriminant analysis (LDA), and naive Bayesian classifier (NBC) techniques. Compared with PCA-LDA algorithms, PCA-NBC techniques together with leave-one-out, cross-validation method provide better discriminative results of normal, adenomatous polyp, and adenocarcinoma gastric tissues, resulting in superior sensitivities of 96.3%, 96.9%, and 96.9%, and specificities of 93%, 100%, and 95.2%, respectively. Therefore, NIR RS associated with multivariate statistical algorithms has the potential for early diagnosis of gastric premalignant lesions and cancer tissues in molecular level. [DOI: 10.1117/1.JBO.18.6.067004]

Keywords: premalignant lesions; adenomatous polyp; gastric cancer; adenocarcinoma; Raman spectroscopy; principal components analysis; linear discriminant analysis; naive Bayesian classifier.

Paper 12798RR received Jan. 21, 2013; revised manuscript received Apr. 25, 2013; accepted for publication Apr. 29, 2013; published online Jun. 4, 2013.

1 Introduction

Currently, gastric cancer is one of the most common malignancies in the world and the second leading cause of cancer death among various cancers.[1,2] It is estimated that a total of 989,600 new gastric cancer cases and 738,000 deaths occurred in 2008 around the world, accounting for 8% of the total cancer cases and 10% of total deaths.[3] More than 70% of new gastric cancer cases and deaths occur in economically developing countries, of which about half occur in China.[3] Ninety percent of all gastric tumors are malignancies, in which gastric adenocarcinoma makes up around 95% of the total number.[4] Although premalignant lesions have a strong tendency to develop into malignancy after being stimulated by some factors for a long time, they still can be cured before they progress to a late stage. It has been demonstrated that the early identification and effective treatment of premalignant lesions can significantly improve the survival rate of patients with gastric cancer, especially those with a high risk of pathologic lesions [e.g., *Helicobacter-pylori* (*Hp*)-infection and gastric intestinal metaplasia (IM)].[5] An early study reported a five-year survival rate of 90% when gastric cancer is detected early and followed by curative endoscopic or surgical therapy.[6] In order to increase the proportion of early cancer detection, the patients with pathologies that are potentially premalignant, such as gastric ulcers, gastric polyps, histological changes of dysplasia, IM, and atrophic gastritis, need to be identified.[7] Certain kinds of polyps, called adenomatous polyps or adenomas, are most likely to become malignancies.[8] The early detection of this pathology through screening and their removal may play an important role in reducing the incidence of gastric cancer and diagnosing cancer at an earlier stage.[1,8] Furthermore, being screened at an appropriate frequency increases the probability of tumor control with immediate treatment and, thus, avoids invasion and metastasis and enables its complete surgical removal.[8,9] Currently, gastroscopy is still the major method for clinical detection of gastric cancer; however, those flat lesions or microlesions are easily overlooked and more difficult to biopsy.[2] Differential diagnosis of protruding lesions based on morphological and/or color characteristics can be difficult, though they have obvious morphological changes that make them easily recognizable under conventional white-light endoscopy.[2] Hence, the ability to reliably differentiate adenomatous polyp from non-neoplastic polyp through visual observation remains limited. On the other hand, biopsy examination remains the gold standard for cancer diagnosis, but it is defective in its invasiveness, long waits, and complicated procedures.[10] Therefore, it is really necessary to develop noninvasive, objective, and sensitive optical diagnostic technologies for detecting premalignant lesions and early cancers during endoscopy. Technologic advancements in conventional white-light endoscopy may help to significantly raise the capacity to diagnose premalignant lesions. As optical adjuncts of gastroscopy, advanced optical spectroscopic and/or imaging techniques do offer many potential advantages, including rapid tissue diagnosis *in situ*, increases in the yield of sampling,

*These authors contributed equally to this work.

Address all correspondence to: Changshui Chen, South China Normal University, Photon & Nano Research Centre for Biosciences, Key Laboratory of Laser Life Science of Ministry of Education, Guangzhou, Guangdong 510631, China. Fax: 0086+20+85211768; E-mail: cschen@aiofm.ac.cn and Hua Mao, Southern Medical University, Zhujiang Hospital, Department of Gastroenterology, Guangzhou, Guangdong 510282, China. E-mail: huam@fimmu.com

0091-3286/2013/$25.00 © 2013 SPIE

identification of suspicious lesions for targeted biopsies, delineation of lesion margins, and reductions in waiting time and the costs and risks associated with unnecessary biopsies.[11] Various types of spectroscopic and/or imaging methods have been extensively applied to the diagnosis of malignant tumors, such as fluorescence spectroscopy and imaging, reflectance spectroscopy, light (elastic) scattering spectroscopy, Raman spectroscopy (RS), optical coherence tomography, chromoendoscopy, narrow-band imaging, immunophotodetection, and, more recently, terahertz imaging and spectroscopy.[12–18]

In particular, RS, which measures inelastic light scattering processes of molecular vibrations and is able to provide more specific spectroscopic features about biomolecular structures and conformations of tissues than those obtained from fluorescence or light (elastic) scattering spectroscopy,[19] has its unique advantages in detecting subtle biomolecular alterations associated with neoplastic transformation, and it may be helpful for us to further understand the molecular mechanism of the cancer development process. RS using near-infrared (NIR) laser light excitation (e.g., 785 nm) is especially attractive as it holds significant advantages over previous spectroscopic techniques in tissue diagnosis, such as deeper penetration into the tissue and far less interference from tissue autofluorescence as compared to ultraviolet/shorter visible light excitation.[20] As a result, NIR RS has been recognized as a valuable tool for *ex vivo* and *in vivo* diagnosis of malignancies in a number of organs, including the stomach.[11,21–23] Teh et al. have achieved high diagnostic accuracies (around 90%) by using NIR RS and a variety of algorithms in distinguishing different pathologic types of gastric tissues (e.g., dysplasia, adenocarcinoma, IM, and *Hp*-infection).[10,24–27] These investigations indicate that significant molecular variations of tissues associated with neoplastic transformation can be exhibited through NIR RS. Bergholt et al.[28–31] have researched the *in vivo* diagnosis of gastric cancer by RS, but the application of clinical Raman endoscopy is still restricted by many technological challenges, such as the efficiency of collecting Raman signals. Although technological progress, including high-sensitivity NIR detectors, filtered fiber-optic probes, and novel spectral analysis techniques,[32–34] will further encourage *in vivo* application of RS, more *ex vivo* research is still required before the final goal of clinical rapid diagnosis based on RS can be completed. For example, investigating the potential of *ex vivo* NIR RS for discriminating adenomatous polyp and adenocarcinoma from normal gastric tissues is valuable. In this study, multivariate statistical techniques are tried to develop effective diagnostic algorithms for the comparison and differentiation of Raman spectra among normal, adenomatous polyp, and adenocarcinoma gastric tissues, including principal component analysis (PCA), linear discriminant analysis (LDA), and naive Bayesian classifier (NBC) algorithms.

2 Materials and Methods

2.1 *Raman Instrumentation*

The instrumentation (Renishaw, InVia + Plus, Confocal Microspectrometer) utilized to measure the Raman spectra of tissues mainly consists of a 785 nm diode laser (maximum output: 300 mW, Renishaw Inc., New Mills, UK), a transmissive imaging spectrograph with an NIR-optimized, back-illuminated, deep-depletion charge-coupled device (CCD) detector, and a single 1200 lines/mm dispersion grating. The original Raman spectra of tissues can be displayed on the computer screen in real time and saved for further analysis. The spectral resolution of the system is 1 cm^{-1} and the laser power used in this study is about 10%. The system acquires each spectrum over the wavenumber range of 700 to 1800 cm^{-1}, which contains the main vibrational signatures of proteins, lipids, carbohydrates, and nucleic acids. The laser power on the surface of the tissue during signal acquisition is ~10 mW, and the diameter of the laser spot at the focus point is about 5 μm. The spectra are collected in backscattering geometry using a microscope equipped with a Leica 20 × objective. In order to obtain a better signal-to-noise ratio, the measurements of all the Raman spectra are accumulated five times within the 5 s of exposure time. The peak frequency of each spectrum and the rapid checking of instrumental performance are calibrated with the silicon phonon line at 520 cm^{-1}. A personal computer controls the Raman spectrometer with a custom-designed software package (Raman Wire 3.2) that triggers on-line data acquisition and analysis.

2.2 *Gastric Tissue Samples*

In this study, the samples of *ex vivo* gastric tissues are collected from 48 gastric patients (29 men and 19 women, with an average age of 61 years). Some of them have undergone the endoscopy biopsies ($n = 37$) for clinically suspicious lesions, and the others pathologically diagnosed with gastric adenocarcinoma have received gastrectomy ($n = 11$) in the Department of Gastroenterology, Zhujiang Hospital, Guangzhou, China. All of those patients have signed an informed consent form to permit the investigative utilization of the tissues. After biopsies or surgical resections, each one of the fresh tissue samples is immediately divided into two portions, and it is ensured that every portion keeps its histological structure complete; one portion is submitted for histopathologic classification and the other is placed in plastic tubes without any treatment and stored on ice until it is sent to the laboratory for Raman measurements. A senior gastrointestinal pathologist, who is unaware of the features of Raman spectra of various tissue types, carries out the analysis of histopathology. The results show that among the 48 relative homogeneous gastric tissue samples selected from a large series of tissue samples with clearly defined pathologies, 22 tissue samples are normal, 13 are adenomatous polyps, and 13 are adenocarcinomas. Each tissue sample is from a different individual; other pathologic tissues or mixture tissue types are excluded in this study. We measure Raman spectra of all tissue samples in air at room temperature immediately after resection. Each tissue sample (approximately 5 × 4 × 3 mm in size) is placed on a glass slide and measured at different regions. After the experiment, each tissue sample is sent to hospital and the tissue diagnosis is confirmed histopathologically. In order to further assess the variability of intrasample, two to three regions of each tissue sample are chosen for multiple measurements ($n > 8$) at random. The random selection is accomplished at the xyz three-dimensional motorized and computer-controlled sample stage. The average spectrum of eight repeated Raman measurements on the same tissue region of each tissue sample is used for spectral analysis. The spectra are examined visually for any obvious outliers that deviate gravely from standard deviation (SD) or are caused by instrumental or man-made measurement errors, and then the obviously aberrant spectra are excluded from analysis. Finally, a total of 105 Raman spectra obtained from 48 relatively homogeneous gastric tissue samples, in which 45 Raman spectra are from normal tissues (i.e., confirmed by histopathology), 27 from premalignant lesions

(i.e., adenomatous polyp as confirmed by histopathology), and 33 from cancer tissues (i.e., adenocarcinoma as confirmed by histopathology), corresponding to two or three average spectra per tissue sample, are eventually applied to the differentiation of gastric tissues in this study.

2.3 *Data Preprocessing*

The original spectra from gastric tissues represent a combination of strong autofluorescence signal, weak Raman signal, and noise. A specially designed software package (Raman Wire 3.2) is used to preprocess these original spectra, including the calibration of wavelength, the calibration of the systematic spectral response, the elimination of cosmic radiation, and the subtraction of the CCD dark-noise. The Raman spectral signal whose noise has been smoothed is collected and removed by a background-correction algorithm named adaptive iteratively reweighted penalized least squares (airPLS), which is implemented in R language.[35,36] Then, each Raman spectrum whose autofluorescence background has been subtracted is also normalized with respect to the total area under the Raman curve from 700 to 1800 cm^{-1}. Thus, the systematic deviation can be minimized and the comparison of the spectral shapes and the relative Raman peak intensities among different types of gastric tissues can be achieved.

2.4 *Multivariate Statistical Analysis*

In this study, PCA, LDA, and NBC techniques are performed on all spectra to elucidate the important spectral features of each tissue class and to compare different tissue classes against each other for gastric tissues identification. PCA is usually employed to extract a set of orthogonal principal components (PCs) to simplify the analysis of data. First, PCA is utilized to reduce the dimensionality of the Raman spectral dataset in our study while still retaining the most important information for tissue diagnosis and characterization. Based on standardized data matrices of Raman spectra, PCA can generate the PCs scores. Each loading vector is related to the original spectrum by PCs score, which represents the weight of that particular component against the basis spectrum.[10,24] A supervised analytical technique named LDA could develop diagnostic algorithms using the PCs scores. LDA determines the discriminant function that maximizes the variances in the data among groups while minimizing the variances between members of the same group.[28] Based on the utilization of a learning algorithm and a probabilistic determination, Bayesian classification rule can relate a small number of spectral metrics to specific histologic entities. The decision rule is not explicitly specified and Bayesian probability might be estimated directly from the using feature of the data distribution.[37,38] One-way analysis of variance (ANOVA)[31,39] is applied for identification of the most significant PCs ($p < 0.05$) in diagnosis for differentiating normal, adenomatous polyp, and adenocarcinoma tissues. These significant PCs scores would be loaded into the model of LDA and NBC to classify those three different kinds of tissues (i.e., normal, adenomatous polyp, and adenocarcinoma tissues). Leave-one-spectrum-out and cross-validation method are used to ensure the validity of the analysis.[24] The results of histopathology serve as the golden standard for assessing the sensitivity and specificity of diagnosing different types of gastric tissues through RS.

3 Results and Discussions

3.1 *Spectral Features*

Figure 1(a) shows the comparison of the average standardized Raman spectra of the normal ($n = 45$, dash-dot line), adenomatous polyp ($n = 27$, solid line), and adenocarcinoma gastric tissues ($n = 33$, short dot line) in the range of 700 to 1800 cm^{-1}. From this figure, it can be seen that the characteristic spectra are similar in shape and intensity among different types of gastric tissues and the spectrum of adenomatous polyp is in the middle of the spectra of normal and adenocarcinoma gastric tissues. Distinct Raman bands are found in those three kinds of tissues at the following peak positions with their respective tentative biomolecular assignments:[9,26,28,31,40–42] 853 cm^{-1} (C-C stretching mode of proline), 936 cm^{-1} (C-C stretching in α-helix conformation of proteins), 1003 cm^{-1} (C-C symmetric stretch ring breathing of phenylalanine), 1032 cm^{-1} (C-H stretching of phenylalanine), 1174 cm^{-1} (C-H bending of

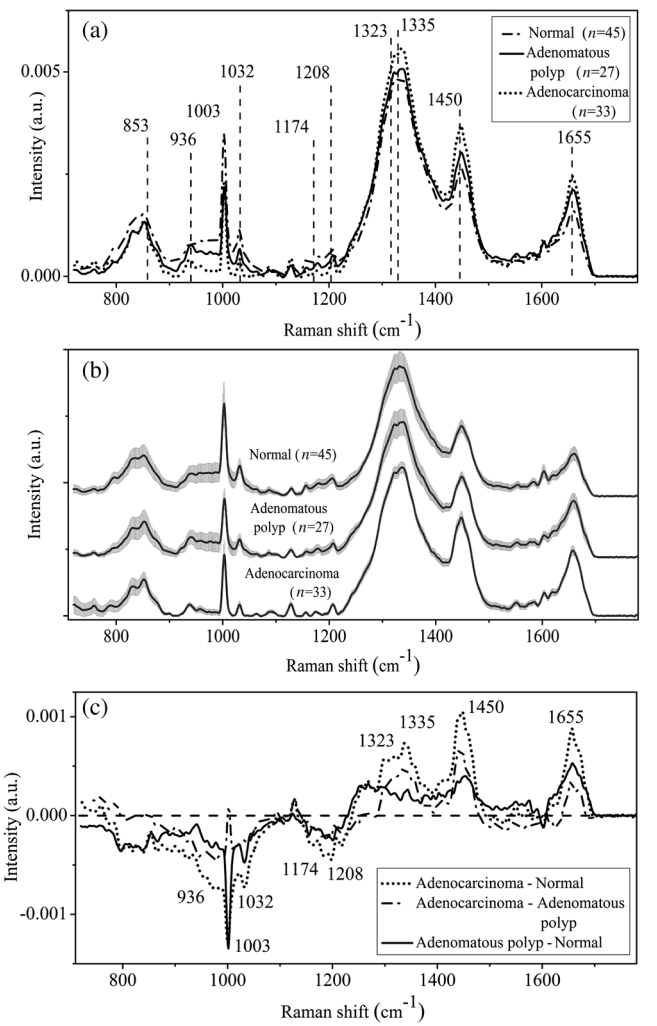

Fig. 1 (a) Comparison of the average standardized Raman spectra with distinct Raman peaks of the normal ($n = 45$, dash-dot line), adenomatous polyp ($n = 27$, solid line), and adenocarcinoma ($n = 33$, short dot line) gastric tissues. (b) Average standardized Raman spectra ± standard deviation of the normal, adenomatous polyp, and adenocarcinoma gastric tissues, respectively. Note that the Raman spectra are vertically shifted for better visualization. (c) Difference spectra are calculated from the average Raman spectra among the three tissue types.

Table 1 Tentative biomolecular assignments[9,26,28,31,40–42] and diagnostic significance (p-value) of the prominent Raman bands identified from all pairwise tissue types based on *post hoc* Fisher's least significant differences (LSD) test.[27,31]

Raman bands (cm^{-1})	Tentative biomolecular assignments	p-values based on the *post hoc* Fisher's LSD test		
		Normal versus adenomatous polyp	Adenomatous polyp versus adenocarcinoma	Normal versus adenocarcinoma
936	ν(C-C) in α-helix conformation of proteins	4.17×10^{-1}	1.74×10^{-1}	1.70×10^{-2}*
1003	ν_s(C – C) ring breathing of phenylalanine	0.10×10^{-2}*	9.31×10^{-1}	0.00*
1032	ν(C – H) of phenylalanine	0.00*	4.5×10^{-2}*	0.00*
1174	δ(C – H) of tyrosine	7.4×10^{-2}	1.48×10^{-1}	0.10×10^{-2}*
1208	ν(C-C_6H_5) of phenylalanine and tryptophan	1.31×10^{-1}	8.70×10^{-2}	0.10×10^{-2}*
1323	Guanine (B, Z marker); CH_3CH_2 wagging and deforming in collagen and purine bases of nucleic acids	4.53×10^{-1}	2.57×10^{-1}	3.90×10^{-2}*
1335	CH_3CH_2 twisting mode of proteins and nucleic acids	4.06×10^{-1}	2.41×10^{-1}	2.90×10^{-2}*
1450	$\delta(CH_2)$ of proteins and lipids	0.30×10^{-2}*	0.00*	0.00*
1655	ν(C = O) of amide I, α-helix in proteins	0.00*	0.10×10^{-2}*	0.00*

Note: ν, stretching mode; ν_s, symmetric stretching mode; δ, bending mode; $*p < 0.05$ (diagnostically significant based on pairwise comparison)

tyrosine), 1208 cm^{-1} (C-C_6H_5 stretching mode of phenylalanine and tryptophan), 1323 cm^{-1} [Guanine (B, Z marker); CH_3CH_2 wagging and deforming in collagen and purine bases of nucleic acids], 1335 cm^{-1} (CH_3CH_2 twisting mode of proteins and nucleic acids), 1450 cm^{-1} (CH_2 bending mode of proteins and lipids), and 1655 cm^{-1} (amide I band C = O stretching mode of proteins, indicating mainly in α-helix conformation). Despite the similarities referred to above, remarkable variances in spectra are observed in the three kinds of tissues. First, compared with the normal tissues, adenomatous polyp tissues display lower intensities at 853, 936, 1003, 1032, 1174, and 1208 cm^{-1}, while displaying higher intensities at 1323, 1335, 1450, and 1655 cm^{-1}. Moreover, compared with the adenocarcinoma tissues, adenomatous polyp tissues reveal lower intensities at 1323, 1335, 1450, and 1655 cm^{-1}, but higher intensities at 936, 1032, 1174, and 1208 cm^{-1}. Second, the spectral bandwidths also display slight variability in the spectral ranges of 880 to 960, 1150 to 1220, and 1640 to 1675 cm^{-1}. In addition, the Raman peak shifts of adenomatous polyp and adenocarcinoma tissues have obvious red shift in the spectral range of 1640 to 1675 cm^{-1} compared with the normal tissues. The average standardized Raman spectra ± SD of the normal, adenomatous polyp, and adenocarcinoma tissues are depicted in Fig. 1(b), respectively. The spectral variabilities of each group are presented in this figure, and the shaded areas in Raman spectra represent the SD of spectral intensities. As shown in Fig. 1(b), spectral variabilities have similar patterns for both classes of normal and adenomatous polyp tissues and appear to be more apparent within the spectral ranges of 790 to 880, 930 to 990, and 1290 to 1400 cm^{-1}, whereas adenocarcinoma tissues have spectral variations in the spectral range of 700 to 860 cm^{-1}.

The difference spectra among the three types of gastric tissues are exhibited more clearly in Fig. 1(c). Overall, the variances of spectral intensities between normal and adenomatous polyps are smaller than those between normal and adenocarcinoma, which illustrates that adenomatous polyp is in the transitional state. From visual inspection, moreover, an inversion in difference spectra is found around 1230 cm^{-1}, which evidences that the intensity of the Raman signal is higher in normal tissues than in pathological tissues (both adenomatous polyp and adenocarcinoma tissues) below 1230 cm^{-1}, and vice versa above 1230 cm^{-1}. In addition, comparison of the Raman signal intensity between adenomatous polyp and adenocarcinoma tissues indicates that the former is notably higher below 1250 cm^{-1}, and vice versa above 1250 cm^{-1}, except for the ranges around 700 to 790, 1000 to 1008, and 1480 to 1640 cm^{-1} that yield an opposite difference. Specially, nine prominent bands in Raman spectra with statistically significant differences among the three tissue types are identified ($p < 0.05$, one-way ANOVA with confidence interval at 95%),[39] including 936, 1003, 1032, 1174, 1208, 1323, 1335, 1450, and 1655 cm^{-1} as revealed in Fig. 1(c), following earlier reported results.[9,10,26–28,31] Table 1 summarizes the tentative biomolecular assignments[9,26,28,31,40–42] of the nine significant Raman bands as well as the corresponding p-values generated from the *post hoc* Fisher's least significant differences (LSD) test[27,31] of all pairwise tissue type comparisons (i.e., normal versus adenomatous polyp, adenomatous polyp versus adenocarcinoma, and normal versus adenocarcinoma). The variances of Raman spectral intensities are probably due to the differences of disease progression. For instance, the intensities of Raman peaks in the vicinity of 853, 936, 1032, 1174, and 1208 cm^{-1} decrease gradually from normal to cancer tissues [Fig. 1(a)], which may be due to the reductions in the percentage of certain biochemical constituents related to early pathological changes, such as proline, tyrosine, tryptophan, and phenylalanine, relative to the total Raman-active components.[26,28,29,40] On the other hand, when normal tissues develop into the adenomatous polyp and adenocarcinoma tissues, there are progressively increased Raman signals, which attribute to proteins and nucleic acids [e.g., 1323, 1335, 1450, and 1655 cm^{-1} in

Fig. 1(a)]. The Raman intensities of those locations in adenomatous polyp and adenocarcinoma tissues enhance markedly, which may be caused by the increase in nucleic acids concentration, amide I of proteins and nucleo-cytoplasmic ratio. All of those are the major characteristics of tissue carcinogenesis.[9,10,14,25,26,43] More specifically, compared with the normal tissues, the increased Raman band intensity and the bandwidth broadening of the 1655 cm^{-1} amide I vibration are highly associated with higher content of α-helix proteins (e.g., histone, the main protein component that makes up the chromatin) in adenomatous polyp and adenocarcinoma tissues.[24–26,28,31] Those discoveries are, in fact, in agreement with the gastric cytological studies for grading malignancy based on the hyperchromatic state of nucleus.[30,43] The alterations of protein-related Raman signals in pathological tissues may suggest the different proteomic activities in the cytoplasm or nucleus and the changes in the extracellular matrix,[20,24–26,42] whereas the overall changes of Raman signals representing lipids and nucleic acids can be associated with an increase of proliferation activity.[20,22,25] The consistency in finding these significant changes of spectral features reinforces the importance of RS for detecting correlative subtle biochemical changes at the molecular level. This provides evidence that the changes of Raman spectral features, which are related to gradual alterations of biochemical constituents and structures in adenomatous polyp and adenocarcinoma tissues, may reflect the development of neoplasia.

3.2 *Diagnosis Potential of RS Using PCA-LDA and PCA-NBC Algorithms*

The diagnostically significant PC loadings computed by PCA are shown in Fig. 2, along with the percentage of spectral variance captured by each of the respective PCs (PC1, PC2, PC3, PC9, PC10, $p < 0.05$, one-way ANOVA with confidence interval at 95%). Obviously, PC1 accounting for the largest Raman spectral variance (~64.04% of the total variance) primarily represents the variations of molecule-specific Raman signals in the major Raman peaks (e.g., at 853, 936, 1003, 1174, 1208, 1323, 1335, 1450, and 1655 cm^{-1}), whereas the successive PCs describe the spectral features that contribute progressively smaller variances (i.e., PC2 ~ 15.36%, PC3 ~ 5.73%, PC9 ~ 0.55%, PC10 ~ 0.43%). According to Fig. 2, the positive and negative band features indicate the correlative direction between PCs and Raman bands. For instance, contributions from 853, 1003, 1174, 1208, and 1450 cm^{-1} are mostly positive

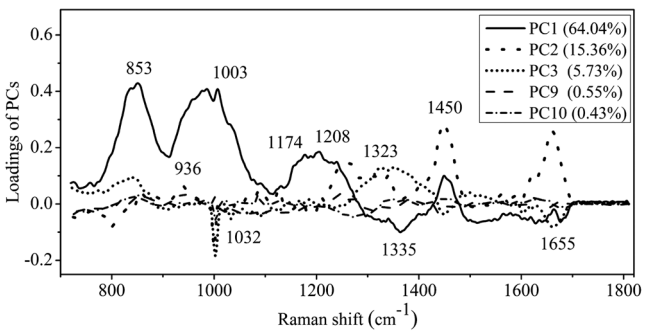

Fig. 2 The diagnostically significant PC loadings (PC1, PC2, PC3, PC9, PC10, $p < 0.05$, one-way ANOVA) capturing a total variance of 86.1% in the full range of the Raman spectrum calculated from Raman spectra of gastric tissues, revealing the diagnostically significant spectral features for tissue classification.

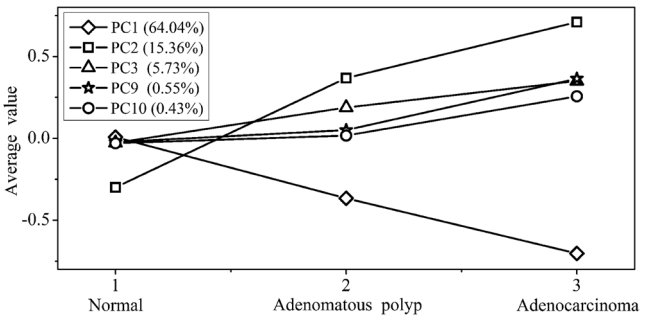

Fig. 3 Broken line graph of average values of the five diagnostically significant PCs (PC1, PC2, PC3, PC9, and PC10), demonstrating the efficacy of PC scores for classification of the normal, adenomatous polyp, and adenocarcinoma gastric tissues.

correlation to PC1, whereas those from 1335 and 1655 cm^{-1} are negatively correlated with PC1. PC loadings essentially reveal the significant Raman spectral features pertinent to proteins, lipids, and nucleic acids within the three different types of gastric tissues, and their features are similar to those of tissue Raman spectra in Fig. 1(a), substantiating that the PCA model has the ability to extract the diagnostic information carried by RS from gastric tissues. Figure 3 displays the estimated average values of the five diagnostically significant PCs for identifying the three different types of gastric tissues. The first three PCs (PC1, PC2, and PC3) account for the maximum variance (~85.13%) of the total variance in the whole range of the Raman spectrum (i.e., 700 to 1800 cm^{-1}); thus they contain the most diagnostically significant Raman features for differentiating normal, adenomatous polyp, and adenocarcinoma tissues. From Fig. 3, it can be seen that the average values of PCs tend to increase gradually with the development of neoplasia, but PC1 exhibits an opposite trend, which demonstrates the efficacy of PC scores in identification of the three different types of gastric tissues. The diagnostically significant PCs together with the corresponding *p*-values calculated from the *post hoc* Fisher's LSD test of all pairwise tissue type comparisons are listed in Table 2. According to these results, different PC scores have different degrees of diagnostic capacity for classifying the tissue types. For instance, PC1, PC3, and PC10 can be used to differentiate

Table 2 Five diagnostically significant PCs of all pairwise tissue types as well as the corresponding *p*-values generated from *post hoc* Fisher's LSD test.

	p-values based on the *post hoc* Fisher's LSD test		
PCs	Normal versus adenomatous polyp	Adenomatous polyp versus adenocarcinoma	Normal versus adenocarcinoma
PC1	7.20×10^{-2}	1.27×10^{-1}	0.00^{*}
PC2	0.00^{*}	$3.8 \times 10^{-2*}$	0.00^{*}
PC3	1.81×10^{-1}	3.46×10^{-1}	$1.40 \times 10^{-2*}$
PC9	5.88×10^{-1}	$3.00 \times 10^{-2*}$	$3.00 \times 10^{-3*}$
PC10	7.22×10^{-1}	8.90×10^{-2}	$2.20 \times 10^{-2*}$

Note: $*p < 0.05$ (diagnostically significant based on pairwise comparison)

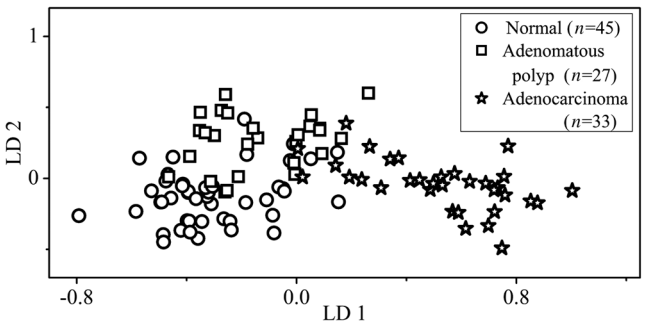

Fig. 4 Scatter plot of the two linear discriminant functions (LD1 versus LD2) based on the PCA-LDA model for discrimination of the normal, adenomatous polyp, and adenocarcinoma gastric tissues.

adenocarcinoma from normal tissues; PC2 is optimal in discriminating the three different types of gastric tissues; and PC9 can be utilized to distinguish adenocarcinoma from adenomatous polyp and normal tissues.

In order to further improve the accuracy of the discrimination, we load the diagnostically significant PC scores into the LDA and NBC model to generate effective diagnostic algorithms for tissue groupings. Figure 4 exhibits the two-dimensional scatter plot of the two linear discriminant (LD) functions (LD1 versus LD2) based on the PCA-LDA model, in which obvious clusterings of the tissue types are largely separated into three diagnostic groups [i.e., normal (circle), adenomatous polyp (square), and adenocarcinoma (pentacle)]; the mixed results may be attributed to the similar constituents within tissues. The discrimination results based on NIR RS using PCA-LDA and PCA-NBC algorithms associated with leave-one-out, cross-validation method are summarized in Table 3. Diagnostic sensitivities of 81.5%, 85.3%, and 100%, and specificities of 86.4%, 100%, and 97.4%, respectively, are achieved by using PCA-LDA algorithms to discriminate the normal, adenomatous polyp, and adenocarcinoma gastric tissues, while PCA-NBC algorithms provide the diagnostic sensitivities of 96.3%, 96.9%, and 96.9%, and specificities of 93%, 100%, and 95.2%, respectively, for identifying those three kinds of tissues. The diagnostic algorithms based on PCA-LDA and PCA-NBC techniques have offered accurate information for gastric tissue classification, further demonstrating a considerable value of RS for tissue diagnosis. Compared with PCA-LDA algorithms, PCA-NBC techniques together with leave-one-out, cross-validation method have realized more effective discriminating capacity for differentiating adenomatous polyp and adenocarcinoma from the normal gastric tissues with a higher sensitivity and specificity.

To find the embodiment of the regularity of pathological development of cancer on Raman spectral characteristic, we explore the Raman spectral properties of normal, adenomatous polyp, and adenocarcinoma gastric tissues in this paper. The results confirm a potential role of RS for revealing the rules of biochemical/biomolecular alterations associated with normal-to-cancer transition in the stomach. However, it is noteworthy that this *ex vivo* study may not truly reflect *in vivo* clinical conditions during endoscopic inspection. The *in vivo* Raman signals may contain a mixture of spectral information from different pathological tissues due to the complexity in the body's internal environment. Nevertheless, the regularity of biochemical/biomolecular changes is fundamentally similar in both *in vivo* and *ex vivo* conditions. For this reason, the results of this *ex vivo* study will provide important experimental basis to the study of *in vivo* real-time diagnosis of premalignant lesions and early gastric cancers based on RS during endoscopy. Despite the great advantages that NIR RS could offer, however, *in vivo* Raman measurements with fiber-optic probes have many technical challenges to overcome because of the weak intensity of the Raman signal from tissue, interference from tissue autofluorescence, and spectral contamination of background fluorescence and Raman signals from the fiber-optic material comprised in the probe.[11,24] Also, the collection times and laser powers are the main limitations for clinical application. Besides, Raman spectral differences are usually subtle between different pathological types of gastric tissues.[10,24] Consequently, it is highly required to develop robust diagnosis algorithms to extract diagnostically important information that is related to biochemical/biomolecular changes associated with neoplastic transformation for effective tissue classification. The effective settlement of these problems will greatly promote the development of an NIR Raman endoscopic system that is a potentially clinically useful tool for the rapid, noninvasive and *in vivo* early diagnosis of gastric premalignant lesions at the molecular level.

4 Conclusions

Early identification of premalignant lesions is an essential factor toward the improvement of clinical outcomes, by developing noninvasive and sensitive optical diagnostic modalities that can assist in guiding physicians for the targeted surveillance and biopsies, thus enhancing the survival rate and the quality

Table 3 Identification results based on NIR Raman spectra of the normal ($n = 45$), adenomatous polyp ($n = 27$), and adenocarcinoma ($n = 33$) gastric tissues utilizing PCA-LDA and PCA-NBC algorithms together with leave-one-out, cross-validation method.

	PCA-LDA classification results			PCA-NBC classification results		
Tissue type	Normal	Adenomatous polyp	Adenocarcinoma	Normal	Adenomatous polyp	Adenocarcinoma
Normal (45)	38	6	1	40	3	2
Premalignant lesions (adenomatous polyp = 27)	5	22	0	1	26	0
Cancer (adenocarcinoma = 33)	0	4	29	1	1	31
Sensitivities (%)	81.5	85.3	100	96.3	96.9	96.9
Specificities (%)	86.4	100	97.4	93.0	100	95.2

of life. As RS has the capacity of obtaining a rich source of diagnostic information about specific biochemical alterations within tissues, it can, therefore, help us further understand the biological processes involved in the transitional period from normal to cancer. Multivariate statistical techniques (e.g., PCA, LDA, and NBC) that utilize the entire Raman spectrum to elucidate the most diagnostically significant spectral features can raise the diagnostic efficiency for tissue classification. In this study, a first attempt toward the investigation of significant differences among the *ex vivo* normal, adenomatous polyp, and adenocarcinoma gastric tissues using NIR RS is described. It is found that NIR RS combined with PCA-NBC techniques can effectively illustrate those biomolecular alterations in biochemical and molecular level, which demonstrates the potential of RS for discriminating the normal gastric tissues, premalignant lesions, and cancer tissues. The categorizing results of adenomatous polyp tissues indicate that the premalignant lesions are diagnostically different from the normal and adenocarcinoma gastric tissues. This phenomenon reflects that the malignant molecular changes have emerged in adenomatous polyp, which is in accordance with what has been expected in the premalignant process. Hence, NIR RS can not only be utilized to differentiate cancer tissues from normal gastric tissues but also provides a potential way for early diagnosis of premalignant lesions. The research of *ex vivo* Raman spectra in this paper also has important clinical reference values in exploring the method of using RS for noninvasive, *in vivo* diagnosis and the screening of early gastric cancer.

Acknowledgments

This work was cofunded by the Key Program of Natural Science Foundation of Guangdong province (Nos. 10251063101000001 and 8251063101000006) and the National Natural Science Foundation of China (No. 60878063)

References

1. C. J. Clark et al., "Current problems in surgery: gastric cancer," *Curr. Probl. Surg.* **43**(8–9), 566–570 (2006).
2. A. Axon, "Symptoms and diagnosis of gastric cancer at early curable stage," *Best Pract. Res. Cl. Ga.* **20**(4), 697–708 (2006).
3. A. Jemal et al., "Global cancer statistics," *CA. Cancer J. Clin.* **61**(2), 69–90 (2011).
4. B. J. Dicken et al., "Gastric adenocarcinoma: review and considerations for future directions," *Ann. Surg.* **241**(1), 27–39 (2005).
5. R. A. Busuttil and A. Boussioutas, "Intestinal metaplasia: a premalignant lesion involved in gastric carcinogenesis," *J. Gastroenterol. Hepatol.* **24**(2), 193–201 (2009).
6. H. Yamamoto and H. Kita, "Endoscopic therapy of early gastric cancer," *Best Pract. Res. Cl. Ga.* **19**(6), 909–926 (2005).
7. J. L. Whiting et al., "The long term results of endoscopic surveillance of premalignant gastric lesions," *Gut.* **50**(3), 378–381 (2002).
8. American Cancer Society, *Colorectal Cancer Facts & Figures 2008–2010*, American Cancer Society, Atlanta (2008).
9. P. C. Lopes et al., "Discriminating adenocarcinoma from normal colonic mucosa through deconvolution of Raman spectra," *J. Biomed. Opt.* **16**(12), 127001 (2011).
10. S. K. Teh et al., "Near-infrared Raman spectroscopy for early diagnosis and typing of adenocarcinoma in the stomach," *Br. J. Surg.* **97**(4), 550–557 (2010).
11. A. Molckovsky et al., "Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps," *Gastrointest. Endosc.* **57**(3), 396–402 (2003).
12. X. Shao, W. Zheng, and Z. Huang, "Polarized near-infrared autofluorescence imaging combined with near-infrared diffuse reflectance imaging for improving colonic cancer detection," *Opt. Express.* **18**(23), 24293–24300 (2010).
13. I. Georgakoudi et al., "Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with Barrett's esophagus," *Gastroenterology.* **120**(7), 1620–1629 (2001).
14. Z. Huang et al., "*In vivo* early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy," *J. Biomed. Opt.* **15**(3), 037017 (2010).
15. X. Shao, W. Zheng, and Z. Huang, "Near-infrared autofluorescence spectroscopy for in vivo identification of hyperplastic and adenomatous polyps in the colon," *Biosens. Bioelectron.* **30**(1), 118–122 (2011).
16. J. C. Taylor et al., "Optical adjuncts for enhanced colonoscopic diagnosis," *Br. J. Surg.* **94**(1), 6–16 (2007).
17. R. S. DaCosta, B. C. Wilson, and N. E. Marcon, "Optical techniques for the endoscopic detection of dysplastic colonic lesions," *Curr. Opin. Gastroenterol.* **21**(1), 70–79 (2005).
18. G. Reese et al., "Using terahertz pulsed imaging (TPI) to identify colonic pathology," in *IEEE 33rd Int. Conf. on Infrared, Millimeter and Terahertz Waves, IRMMW-THz*, California Institute of Technology, Pasadena, California (2008).
19. R. S. Dacosta, B. C. Wilson, and N. E. Marcon, "New optical technologies for earlier endoscopic diagnosis of premalignant gastrointestinal lesions," *J. Gastroenterol. Hepatol.* **17**(Suppl.), S85–S104 (2002).
20. Z. Huang et al., "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer.* **107**(6), 1047–1052 (2003).
21. P. R. T. Jess et al., "Early detection of cervical neoplasia by Raman spectroscopy," *Int. J. Cancer.* **121**(12), 2723–2728 (2007).
22. S. K. Teh et al., "Spectroscopic diagnosis of laryngeal carcinoma using near-infrared Raman spectroscopy and random recursive partitioning ensemble techniques," *Analyst.* **134**(6), 1232–1239 (2009).
23. T. Kawabata et al., "Optical diagnosis of gastric cancer using near-infrared multichannel Raman spectroscopy with a 1064-nm excitation wavelength," *J. Gastroenterol.* **43**(4), 283–290 (2008).
24. S. K. Teh et al., "Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue," *Br. J. Cancer* **98**(2), 457–465 (2008).
25. S. K. Teh et al., "Diagnosis of gastric cancer using near-infrared Raman spectroscopy and classification and regression tree techniques," *J. Biomed. Opt.* **13**(3), 034013 (2008).
26. S. K. Teh et al., "Near-infrared Raman spectroscopy for gastric precancer diagnosis," *J. Raman Spectrosc.* **40**(8), 908–914 (2009).
27. S. K. Teh et al., "Near-infrared Raman spectroscopy for optical diagnosis in the stomach: identification of Helicobacter-pylori infection and intestinal metaplasia," *Int. J. Cancer* **126**(8), 1920–1927 (2010).
28. M. S. Bergholt et al., "Combining near-infrared-excited autofluorescence and Raman spectroscopy improves in vivo diagnosis of gastric cancer," *Biosens. Bioelectron.* **26**(10), 4104–4110 (2011).
29. M. S. Bergholt et al., "Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach," *Analyst* **135**(12), 3162–3168 (2010).
30. M. S. Bergholt et al., "*In vivo* diagnosis of gastric cancer using Raman endoscopy and ant colony optimization techniques," *Int. J. Cancer.* **128**(11), 2673–2680 (2011).
31. M. S. Bergholt et al., "Characterizing variability in in vivo Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection," *J. Biomed. Opt.* **16**(3), 037003 (2011).
32. E. B. Hanlon et al., "Prospects for *in vivo* Raman spectroscopy," *Phys. Med. Biol.* **45**(2), R1–R59 (2000).
33. W. Petrich, "Mid-infrared and Raman spectroscopy for medical diagnostics," *Appl. Spectrosc. Rev.* **36**(2), 181–237 (2001).
34. S. Duraipandian et al., "Real-time Raman spectroscopy for *in vivo*, online gastric cancer diagnosis during clinical endoscopic examination," *J. Biomed. Opt.* **17**(8), 81418 (2012).
35. Z. M. Zhang et al., "An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy," *J. Raman Spectrosc.* **41**(6), 659–669 (2010).
36. Z. M. Zhang, S. Chen, and Y. Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst* **135**(5), 1138–1146 (2010).
37. C. Krafft et al., "Disease recognition by infrared and Raman spectroscopy," *J. Biophoton.* **2**(1–2), 13–28 (2009).

38. R. Bhargava et al., "High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data," *Biochimica et Biophysica Acta (BBA)–Biomembranes.* **1758**(7), 830–845 (2006).
39. R. Bender and S. Lange, "Adjusting for multiple testing-when and how?," *J. Clin. Epidemiol.* **54**(4), 343–349 (2001).
40. R. E. Kast et al., "Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model," *Biopolymers* **89**(3), 235–241 (2008).
41. X. L. Yan et al., "Raman spectra of single cell from gastrointestinal cancer patients," *World J. Gastroenterol.* **11**(21), 3290–3292 (2005).
42. A. Mahadevan-Jansen and R. Richards-Kortum, "Raman spectroscopy for the detection of cancers and precancers," *J. Biomed. Opt.* **1**(1), 31–70 (1996).
43. J. H. Hughes et al., "Cytologic criteria for the brush diagnosis of gastric adenocarcinoma," *Cancer* **84**(5), 289–294 (1998).