

# Analyst

Accepted Manuscript



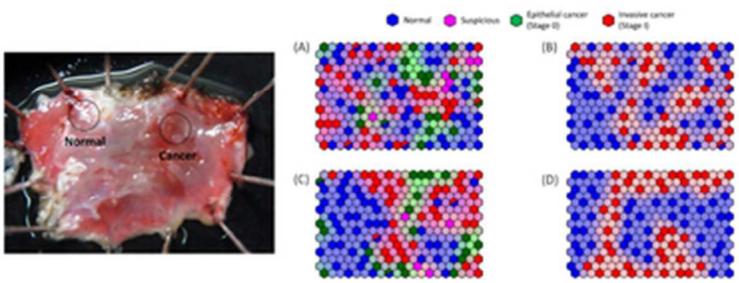
This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



32x13mm (300 x 300 DPI)

1  
2  
3  
4  
5 We evaluated the potential of Raman spectroscopy for the diagnosis of  
6 early stage esophageal cancer with chemometric techniques.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# Diagnosis of early-stage esophageal cancer by Raman spectroscopy and chemometric techniques

Mika Ishigaki<sup>a, b\*</sup>, Yasuhiro Maeda<sup>a, c</sup>, Akinori Taketani<sup>a</sup>, Bibin B Andriana<sup>a</sup>,  
Ryu Ishihara<sup>d</sup>, Kanet Wongravee<sup>e</sup>, Yukihiko Ozaki<sup>b</sup>, and Hidetoshi Sato<sup>a\*</sup>

- a. Department of Biomedical Chemistry, School of Science and Technology, Kwansai Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan
- b. Department of Chemistry, School of Science and Technology, Kwansai Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan
- c. Riken Center for Advanced Photonics, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
- d. Osaka Medical Center for Cancer and Cardiovascular Diseases, 1-3-3 Nakamichi, Higashinari-ku, Osaka, 537-8511, Japan
- e. Sensor Research Unit (SRU), Department of Chemistry, Chulalongkorn University, 254 Phayathai Road, Patumwan, Bangkok, 10330, Thailand

\* Authors to whom correspondence should be sent.

## ABSTRACT

Esophageal cancer is a disease with high mortality. In order to improve the 5-year survival rate after cancer treatment, it is important to develop a method for early detection of the cancer and for therapy support. There is increasing evidence that Raman spectroscopy, in combination with chemometric analysis, is a powerful technique for discriminating pre-cancerous and cancerous biochemical changes. In the present study, we used Raman spectroscopy to examine early-stage (stages 0 and I) esophageal cancer samples *ex vivo*. Comparison between the Raman spectra of cancerous and normal samples using a *t*-test showed decreased concentrations of glycogen, collagen, and tryptophan in cancerous tissue. Partial least squares regression (PLSR) analysis and self-organization maps (SOMs) discriminated the datasets of cancer and normal samples into two groups, but there was a relatively large overlap between them. Linear discriminant analysis (LDA) based on Raman bands found in the *t*-test was able to predict the tissue types with 81.0% sensitivity and 94.0% specificity.

**Keyword:** Raman spectroscopy, early esophageal cancer, chemometrics, discriminant analysis

## 1. INTRODUCTION

Esophageal cancer has high mortality because of its tendency to spread to other tissues or organs.<sup>1-3</sup> Surgery may also contribute to the high mortality because many vital organs surround the esophagus. Recent progress in diagnostic techniques and medical treatments for esophageal cancer has improved

1  
2  
3  
4 the 5-year survival rate after treatment. The Japanese Association of Clinical Cancer Centers reports  
5 that the 5-year survival rates after treatment for stage 0 and I cancers are about 80% and 40%,  
6 respectively.<sup>4</sup> This indicates that the detection of the esophageal cancer at an early stage is essential for  
7 better outcome of the cancer therapy.  
8

9  
10 The breakthrough in medical technology with endoscopic techniques such as narrow band imaging  
11 (NBI) and auto fluorescence imaging (AFI) is responsible for the early detection and rapid cure of  
12 esophageal cancer. These endoscopic techniques improve color contrast and brightness and emphasize  
13 edges, making the shape and pattern of the tissue, such as the capillary blood vessels, clearer. These  
14 visual cues are useful for detecting abnormal tissue during the endoscopy.<sup>5</sup> Another popular technique  
15 to detect the cancer lesion during the endoscopic medical check-up is the indigo carmine dye method.<sup>6,7</sup>  
16 In this method, the deep-blue indigo carmine dye is instilled into the digestive tract during the  
17 endoscopic observation to make the asperity of the mucosal surface visually clear. This method is very  
18 useful to detect abnormal tissues in the wide ductal epithelial. These equipment and methods are helpful  
19 to detect the cancer lesion at an early stage and improve diagnosis.<sup>8,9</sup> Despite their usefulness, the present  
20 methods of cancer detection depend on visual pattern recognition and rely on the skills of each medical  
21 doctor. These skills are currently quite difficult to replace with computerized automated diagnosis. The  
22 same scenario applies to histopathological diagnosis, where cancer detection depends on morphological  
23 analysis by a doctor using a microscope. Hence, the reliability and precision of cancer diagnosis largely  
24 depend on the skills and experience of each doctor.  
25

26  
27 Since the morphological changes should take place after the changes at the molecular level, such as  
28 concentration changes, structural changes, and residue changes due to cancer metabolism, the detection  
29 of abnormal molecular alterations in tissues may be essential for cancer detection at an early stage.  
30 Histopathological analysis including immunostaining can reliably detect molecular changes but it is  
31 invasive and time-consuming. In order to save the doctor's time and labor, an on-time label-free method  
32 for cancer diagnosis is in demand. Raman spectroscopy is one of the most powerful techniques for this  
33 purpose.  
34

35  
36 Usually Raman spectroscopy does not request any sample preparation. Furthermore, strong water  
37 bands in a Raman spectrum do not overlap with the fingerprint region, so that Raman spectroscopy suits  
38 biological applications better than IR spectroscopy. Raman spectroscopy provides molecular  
39 information noninvasively and *in situ* without labeling, and thus, many papers on Raman applications  
40 to biological and medical subjects have been reported.<sup>10-24</sup> In the field of its application to cancer  
41 diagnosis in fact, a number of research groups have been involved in exploring the feasibility for cancer  
42 diagnosis in various organs (breast,<sup>12,13</sup> colon,<sup>14</sup> cervix,<sup>15,16</sup> esophagus,<sup>17-19</sup> stomach,<sup>20</sup> oral,<sup>21</sup> skin,<sup>22</sup>  
43 brain<sup>23</sup>). For instance, Haka et al. diagnosed breast cancer based on Raman spectra from fresh-frozen *in*  
44 *vitro* samples with sensitivity of 83% and specificity of 93%.<sup>13</sup> Zheng et al. proved the diagnostic power  
45 of Raman spectroscopy over colorectal cancer by the analysis of single live epithelial cells, and their  
46 sensitivity and specificity were 86.3% and 86.3%, respectively.<sup>14</sup> Especially, Huang et al. succeeded in  
47 establishing an automated on-line diagnostic framework for *in vivo* cancer detection at endoscopy with  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Raman spectroscopy. Their diagnostic sensitivity and specificity were 90.0% and 73.3%, respectively.<sup>24</sup>

4  
5 The purpose of this study is to evaluate the potential of Raman spectroscopy for the detection of  
6 early-stage esophageal cancer (stage 0 and I). A homemade portable Raman system with a miniaturized  
7 Raman probe developed by a member of our group<sup>25</sup> was used for *ex vivo* measurements of fresh human  
8 tissues before the histopathological examination in the hospital; the spectroscopic data were analyzed  
9 together with the histological data.  
10

11  
12 Despite the strong autofluorescence background in the Raman spectra, bands characteristic for living  
13 organisms were observed, with only slight spectral differences between cancerous and normal tissues.  
14 Spectral noise hindered data analysis and single chemometric technique was not sufficient to classify  
15 the tissue types accurately. Therefore, we performed partial least square regression (PLSR) analysis and  
16 linear discriminant analysis (LDA) on only the significant wavenumbers assessed by the *t*-test to have  
17 statistically different Raman signal intensity. Furthermore, self-organization maps (SOMs) were also  
18 found suitable for the analysis of the present data, since it is one of the artificial algorithms for neural  
19 networks whose learning process for the pattern recognition resembles the one that is found in the brain.  
20 Particularly, they show their power in pattern recognition involving noisy signals.<sup>26</sup> The results from the  
21 *t*-test, PLSR, and SOMs showed that the concentrations of glycogen, collagen, and tryptophan decrease  
22 in cancerous lesions. The diagnostic sensitivity and specificity calculated through LDA with significant  
23 wavenumbers assessed by the *t*-test amounts to 81.5% and 94.0%, respectively.  
24

25  
26 The results indicate the possibility of detecting early-stage esophageal cancer based on molecular  
27 information rather than morphological features. Diagnosis with Raman spectroscopy leads to early  
28 detection of the pathological changes that take place before the morphological changes and can result in  
29 a better outcome of the cancer therapy. Furthermore, it holds potential for assessment of the effects of  
30 chemical treatment. Spectral diagnosis pioneers a new diagnostic technique that is less invasive, label-  
31 free, and with less burden for the patients.  
32

## 33 34 35 36 37 38 39 40 41 42 **2. MATERIALS AND METHODS**

### 43 44 **2.1 Esophageal cancer**

45 In Japan, the squamous cell cancer accounts for more than 90% of total esophageal cancer.<sup>27</sup> In the  
46 present study, only squamous esophageal cancer was examined. The “Guidelines for Diagnosis and  
47 Treatment of Carcinoma of the Esophagus” of the Japan Esophageal Society define the early stages of  
48 carcinoma progression.<sup>28,29</sup> In stage 0 (epithelial cancer), the carcinoma stays in the mucosa, and lymph  
49 node metastasis and distant metastasis cannot be observed. In stage I (invasive cancer), the carcinoma  
50 either has infiltrated the submucosa or has metastasized to the neighboring lymph nodes. In these early  
51 stages, the cancerous tissue can be removed by endoscopic mucosal resection (EMR).  
52

53 Iodine dye is often used for the endoscopic diagnosis of esophageal cancer. When the iodine dye  
54 is spread onto a gastrointestinal area, it reacts with glycogen and turns the cells brown. Since the  
55 cancerous tissue has reduced glycogen concentration as compared to normal tissue, a normal region  
56 turns brown while a cancerous region remains white.<sup>30-37</sup> The suspicious tissues were resected and  
57  
58  
59  
60

1  
2  
3 immediately sent for Raman measurements, which were performed within 1 h after the resection. The  
4 iodine dye was neutralized by spraying with Sodium Thiosulphate Solution (STSS, Detoxol®)<sup>38</sup> and  
5 resected sample was put into saline. It was also verified that iodine and STSS signals were not be  
6 detected from the resected tissue by Raman spectroscopy. Fifteen samples from different patients were  
7 obtained from patients with stage 0 and I esophageal cancers that had not been chemically treated before  
8 the resection. A photographic image of a resected tissue is shown in Figure 1. As the tissues always  
9 included both normal and carcinoma areas, the spectral measurements were performed at three to five  
10 points in the normal and carcinoma tissue areas on the side of esophageal lumen. The sample size was  
11 several centimeter square and the thickness was about 1 mm. After the Raman measurements, the tissues  
12 were immediately fixed with formalin and sent for histopathological examination. Pathologist confirmed  
13 that there were no degenerations of the tissue caused by measurement time and laser irradiation. The  
14 present study was approved by an ethical committee of Kwansai Gakuin University and Osaka Medical  
15 Center for Cancer and Cardiovascular Diseases, and the sample collection was conducted with the  
16 informed consent of the patients.  
17  
18  
19  
20  
21  
22  
23  
24

## 25 26 27 **2.2 Raman measurements**

28 The portable Raman system was set up in the room next door to the operation room. The system consisted  
29 of a 785nm diode laser (FSL-785-400MM, EnWave Optronics, Inc., USA), spectrometer (F=1.8, Kaiser  
30 Optical System) and CCD detector (DU401-BR-DD, ANDOR). A micro-Raman probe was coupled to the  
31 Raman system.<sup>25</sup> The diameter of the laser spot was about 1 mm. The probe was held at a distance of 1  
32 mm from the surface of the tissue. The excitation laser power was 60 mW at the sample point and exposure  
33 time was 180 s (3×60 s). The background noise due to the Raman probe and the Raman system itself was  
34 subtracted and the spectrum was preprocessed with Savitzky–Golay smoothing (convolution width of 13).  
35 The autofluorescence background of the sample was removed using a 5th-order polynomial fitting.  
36  
37  
38  
39  
40  
41

## 42 **2.3 Multivariate analysis**

43 The spectral intensity was normalized with a standard band at 1005 cm<sup>-1</sup> because it was easy to  
44 distinguish the sharp features of intensity at that wavenumber due to phenylalanine. The processed  
45 spectra were labeled cancerous or normal after confirmation from histopathological examinations. The  
46 spectra were analyzed using *t*-test, PLSR, SOMs, and LDA to extract the slight molecular changes in  
47 order to obtain better discrimination between the different tissues. The calculations were performed with  
48 chemometrics software Unscrambler (CAMO, USA) and MATLAB (Mathworks Inc., USA).  
49  
50  
51  
52  
53  
54

## 55 **3. RESULTS AND DISCUSSION**

56 Figure 2 shows averaged Raman spectra in the region from 600 to 1800 cm<sup>-1</sup> for the carcinoma (Fig.  
57 2a; n = 73: including stage I (n=42), stage 0 (n=25), and suspicious lesion (n=6)) and normal (Fig. 2b;  
58 n = 50) tissues, as well as their subtracted spectra (cancerous tissue spectra minus normal tissue spectra;  
59 Fig. 2c), with ±1 standard error. Histopathological examination could not definitely diagnose about six  
60

1  
2  
3  
4 data whether they were cancer or not. Purandare et al developed the arguments to identify the  
5 intraepithelial neoplasia of cervical tissues that were more likely to progress cancer using IR  
6 spectroscopy.<sup>39</sup> In the present study, whether the tissues in so-called gray zone will progress or regress  
7 was not discussed. We defined them as “suspicious lesion” in our study and they were included in  
8 carcinoma dataset. Although the spectra had strong background due to autofluorescence, it was possible  
9 to observe some bands characteristic of biological tissues, even before data preprocessing. Some  
10 prominent Raman bands were assigned as follows: a symmetric ring-breathing vibration from  
11 tryptophan at 755 cm<sup>-1</sup>; a contribution of proline, tryptophan, and/or glycogen near 855 cm<sup>-1</sup>; and a  
12 symmetric ring-breathing vibration from phenylalanine at 1005 cm<sup>-1</sup>.<sup>40-43</sup> Bands near 1246 cm<sup>-1</sup> are  
13 assigned to amide III and/or C=C-H deformation modes. Bands at 1450 and 1665 cm<sup>-1</sup> are due to C-H  
14 deformation, and amide I and/or C=C stretching modes, respectively. In the subtracted spectra, slight  
15 differences between cancerous and normal tissues were observed. It is assumed that the five negative  
16 bands in the 849–1037 cm<sup>-1</sup> region are characteristic of glycogen and collagen.<sup>40,41</sup> The band at 810 cm<sup>-1</sup>  
17 may originate from the O-P-O stretching mode of RNA.<sup>40</sup> Bands at 1366 and 1627 cm<sup>-1</sup> can be assigned  
18 to tryptophan.<sup>40</sup>

19  
20 We performed unpaired two-sided Student’s *t*-test in order to assess the significant differences (*p*  
21 < 0.05) between the control (normal tissue) and cancer spectra. Six negative bands among the  
22 subtraction spectra were found to be significantly different, as shown in Figure 3. They seem to be the  
23 characteristic bands of glycogen, collagen, and tryptophan. This indicates a decrease in the  
24 concentrations of these substances in the cancerous tissues. We selected these bands as possible  
25 candidates for discrimination factors and used them in the PLSR analysis to construct a discrimination  
26 model. The PLSR score plot is shown in Figure 4A and the score average for each tissue type, with the  
27 corresponding standard errors, are shown in Figure 4B. These figures show that the cancerous (invasive  
28 cancer, epithelial cancer, and suspicious lesion) and normal tissues can be discriminated by the six  
29 negative bands, which originate from glycogen, collagen and tryptophan.

30  
31 SOMs are a powerful tool to visualize data and the neural network statistical model may reveal the  
32 underlying patterns hidden in the obtained datasets.<sup>44-47</sup> In practice, an initial map was built and  
33 represented by a two-dimensional hexagon. Each map unit contains weight vectors randomly generated  
34 from a uniform distribution between the maximum and minimum values of variables in the data.  
35 Therefore, a number of layers in each map will be equal to the number of variables present in the data.  
36 In this study, a map with  $P \times Q = 15 \times 20$  map units, which is approximately 2.5 times the number of  
37 samples in the dataset, was generated with a layer of weights equal to *J* variables. During the  
38 calculations, each sample is projected on each map and the Euclidean distance between each map unit  
39 and the sample is calculated. For each sample, the unit with the lowest distance will be assigned as the  
40 best map unit (BMU). The BMU and its neighbor units are updated to become similar to the sample.  
41 This process is repeated until all samples have been assigned to a unit. Then, the whole procedure will  
42 be iterated for 10000 times. After the learning process is achieved, samples that have similar  
43 characteristics, based on class membership and similarity in measured variables, should be assigned to

1  
2  
3  
4 similar regions on the map. To visualize samples on the map, each group of samples is assigned a  
5 different color. A solid color represents an assigned sample and a pale color represents a neighbor unit.  
6 This algorithm is a neural network because the map develops and adjusts itself to fit with the data.  
7

8 The SOMs were trained using the spectral data with all the wavenumbers and with only the  
9 wavenumbers selected from the *t*-test assessment for visualization in order to investigate the variation  
10 in Raman spectra obtained from normal and cancerous tissues. The SOMs (Figure 5A-D) showed  
11 poorly distinct regions when using all the wavenumbers, whereas good separation was observed from  
12 when only the wavenumber selected from the *t*-test assessment was used. This suggests that the  
13 reduction in the number of variables had a significant impact on the discrimination between the tissues  
14 through their chemical content. However, only a small separation between states of cancer (invasive  
15 cancer, epithelial cancer, and suspicious lesion) was observed. This observation is in good agreement  
16 with the PLSR results shown in Figure 4A. The results obtained from SOMs also indicate that glycogen,  
17 collagen, and tryptophan are the key molecular factors for discriminating the tissue types.  
18

19 LDA was also performed in order to validate the discrimination between the two tissue types with  
20 the six selected band regions, using the leave-one-out cross validation (LOOCV) technique. In this  
21 validation, a sample was left out as a validated sample, and the LDA model was built using the  
22 remaining samples. Then, the generated model was used to predict the class of the validated sample.  
23 This process was repeated until all spectra were classified. The classification was performed on the  
24 data containing 92 samples from only normal (n=50) and invasive cancer (n=42, stage I) tissues,  
25 excluding stage 0 (n=25) and suspicious lesion (n=6). To evaluate the performance of the selected  
26 bands, the fraction of true positive (cancer sample predicted as cancerous tissue) and the false positives  
27 (normal sample predicted as cancerous tissue) were calculated (Table 1). These statistics can be  
28 summarized graphically in a ROC plot as shown in Figure 6, which is obtained by plotting the  
29 proportion of true positive (TPs) against the proportion of false positive (FPs).<sup>48</sup> These two  
30 characteristics are known as the sensitivity and the complement of specificity. To simplify the  
31 interpretation from ROC curve, the area under the ROC curve (AUC) was calculated to visually  
32 summarize the performance of the LDA model. A perfect ROC curve, which means perfect  
33 classification of all samples, has an AUC of 1. A higher value of AUC corresponds to a better  
34 classification. As shown in Figure 6, a good ROC curve with an AUC equal to 0.9410 was obtained  
35 when using a LDA classifier with the LOOCV technique. Furthermore, Table 1 shows the contingency  
36 table of classification. The sensitivity and specificity of the present model were 81.0% and 94.0%,  
37 respectively. This suggests that the selected markers correspond to real differences between normal  
38 and cancerous tissues.  
39

40 The present results suggest that the six significant bands can be used as marker bands for the Raman  
41 diagnosis of early-stage esophageal cancer. They are assigned to proline in collagen, tryptophan, and  
42 glycogen. Reduction in glycogen levels is already used for identifying cancerous regions through the  
43 iodine dye test during the endoscopy, and the present result is consistent with the basic procedure of  
44 the diagnosis. If the sign of glycogen decrease could be detected initially with the noninvasive Raman  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 spectroscopy rather than with the iodine dye staining procedure, it would be even more attractive for  
4 patients because the invasive staining method would be replaced with a less invasive and more  
5 comfortable diagnostic procedure.  
6  
7

8 A few research groups have reported the reduction of collagen concentration in esophageal and  
9 stomach cancers.<sup>49-51</sup> It is known that the overexpression of proteolytic enzymes such as matrix  
10 metalloproteinase (MMP) is observed in tumor cells.<sup>52-54</sup> MMPs degrade many kinds of extracellular  
11 matrix proteins, and this correlates with the progression and invasion of carcinoma.<sup>55-57</sup> Therefore, our  
12 results show that the decrease in collagen concentration might be a natural consequence of the  
13 mechanism underlying cancer growth.  
14  
15

16 Furthermore, many researchers have pointed out the abnormal metabolism of tryptophan in cancer.  
17 Tryptophan is mainly metabolized by the kynurenine pathway and the rate-limiting enzyme of this  
18 pathway is indoleamine-2, 3-dioxygenase (IDO).<sup>58</sup> It has been reported that a large proportion of primary  
19 cancer cell overexpress IDO and that the enhanced metabolism in the cancerous region leads to  
20 tryptophan depletion and the accumulation of its metabolic products.<sup>59</sup> Some of those metabolic products  
21 inhibit T cell mobilization and activation, contributing to the immunosuppressive effect and resulting in  
22 more extensive disease.<sup>60-62</sup> Once again, it is remarkable that our results, in this case the tryptophan  
23 decrease, correlate well with similar facts found in the literature.  
24  
25

26 In summary, the marker bands due to glycogen, collagen, and tryptophan found in this study are  
27 adequate for detecting esophageal cancer, proving that Raman spectroscopy is a practical tool for  
28 noninvasive, label-free endoscopic diagnosis.  
29  
30

#### 31 32 33 34 35 36 **4. CONCLUSION**

37 *Ex vivo* Raman spectra of fresh human esophageal cancer tissues were successfully measured using  
38 portable Raman system. The cancerous tissues were typically of early stage cancer (stages 0 and I). We  
39 found six marker bands assessed by *t*-test, assignable to glycogen, collagen, and tryptophan, which  
40 presented decreased band intensity in the cancerous tissues. The PLSR and SOM analysis demonstrated  
41 that these marker bands are useful to distinguish the cancerous tissues from the normal ones. The  
42 validation result using LDA based on those marker bands provided sensitivity and specificity of 81.0%  
43 and 94.0%, respectively. It is remarkable that the present analysis performed only on cancerous tissues  
44 at an early stage and using a portable Raman system under clinical conditions provided such promising  
45 results. Thus, the present study demonstrates that Raman spectroscopy is not only able to detect early-  
46 stage esophageal cancer but is also ready for practical application to *in situ* cancer diagnosis.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 [1]. F. Yousef, et al. *Am. J. Epidemiol.*, 2008, **168**, 237-249.  
5  
6 [2]. C. Bosetti, et al. *Int. J. Cancer*, 2008, **122**, 1118-1129.  
7  
8 [3]. A. P. Polednak, *Int. J. Cancer*, 2003, **105**, 98-100.  
9  
10 [4]. Survival statistics of Japanese association of Clinical Cancer Center by Kaplan-Meier method,  
11 Japanese Association of Clinical Cancer Centers, [https://kapweb.chiba-cancer-](https://kapweb.chiba-cancer-registry.org/web/general/top.aspx)  
12 registry.org/web/general/top.aspx  
13  
14 [5]. L. M. W. K. Song and K. Chen, *Best Pract. Res. Cl. Ga.*, 2005, **19**, 833-856.  
15  
16 [6]. M. Mori, et al. *Am. J. Gastroenterol.*, 1993, **88**, 701-705.  
17  
18 [7]. M. B. Fennerty, *Gastrointest. Endosc. Clin. North Am.*, 1994, **4**, 297-311.  
19  
20 [8]. N. Uedo, et al. *Gastrointest. Endosc.*, 2005, **62**, 521-528.  
21  
22 [9]. M. A. Kara, et al. *Endoscopy*, 2005, **37**, 929-936.  
23  
24 [10]. Y. Ozaki, and K. Iriyama, *Laser Light Scattering Spectroscopy of Biological Objects*, Elsevier Science  
25 Ltd., 1987, "Potential of Raman spectroscopy in medical science."  
26  
27 [11]. Carey, P. *Biochemical applications of Raman and resonance Raman spectroscopies*. Elsevier, 2012.  
28  
29 [12]. M. D. Keller, et al. *J. Biomed. Opt.*, 2011, **16**, 077006-077006.  
30  
31 [13]. A. S. Haka, et al. *J. Biomed. Opt.*, 2009, **14**, 054023-054023.  
32  
33 [14]. F. Zheng, Y. Qin and K. Chen. *J. Biomed. Opt.*, 2007, **12**, 034002-034002.  
34  
35 [15]. L. E. Kamemoto, et al. *Appl. Spectrosc.*, 2010, **64**, 255-261.  
36  
37 [16]. E. M. Kanter, et al. *J. Biophotonics*, 2009, **2**, 81.  
38  
39 [17]. M. S. Bergholt, et al. *Technol. Cancer Res. T.*, 2011, **10**, 103-112.  
40  
41 [18]. J. Hutchings, et al. *J. Biomed. Opt.*, 2010, **15**, 066015-066015.  
42  
43 [19]. C. Kendall, et al. *J. Pathol.*, 2003, **200**, 602-609.  
44  
45 [20]. M. S. Bergholt, et al. *J. Biomed. Opt.*, 2011, **16**, 037003-037003.  
46  
47 [21]. S. P. Singh, et al. *J. Biomed. Opt.*, 2012, **17**, 1050021-1050029.  
48  
49 [22]. C. A. Lieber, et al. *J. Biomed. Opt.*, 2008, **13**, 024013-024013.  
50  
51 [23]. K. Gajjar, et al. *Anal. Methods*, 2013, **5**, 89-102.  
52  
53 [24]. S. Duraipandian, et al. *J. Biomed. Opt.*, 2012, **17**, 081418-081418.  
54  
55 [25]. Y. Komachi, et al. *Appl. Opt.*, 2005, **44**, 4722-4732.  
56  
57 [26]. T. Kohonen, *Proceedings of the IEEE*, 1990, **78**, 1464-1480.  
58  
59 [27]. Y. Tachimori, et al. *Esophagus*, 2014, **11**, 1-20.  
60  
[28]. H. Kuwano, et al. *Esophagus*, 2015, **12**, 1-30.  
[29]. H. Kuwano, et al. *Esophagus*, 2008, **5**, 61-73.  
[30]. R. Voegeli, *ORL*, 1966, **28**, 230-239.  
[31]. S. Toriie, et al. *Endoscopy*, 1975, **7**, 75-79.  
[32]. A. Mandard, et al. *Endoscopy*, 1980, **12**, 51-57.  
[33]. A. Papazian, et al. *Gastroen. Clin. Biol.*, 1985, **9**, 16-22.  
[34]. M. Endo, T. Takemoto and H. Shirakabe, *Semin. Surg. Oncol.*, John Wiley & Sons, Inc., 1986, **2**, No. 3.

- 1  
2  
3  
4 [35]. M. Mori, et al. *Am. J. Gastroenterol.*, 1993, **88**, 701-705.
- 5 [36]. B. J. Nothmann, J. R. Wright and M. M. Schuster, *Am. J. Dig. Dis.*, 1972, **17**, 919-924.
- 6 [37]. H. Shiozaki, et al. *Cancer*, 1990, **66**, 2068-2071.
- 7 [38]. H. Kameyama, et al. *Digest. Endosc.*, 1994, **6**, 181-186.
- 8 [39]. N. C. Purandare, et al. *Anal. Methods*, 2014, **6**, 4576-4584.
- 9 [40]. Z. Movasaghi, S. Rehman, and I. U. Rehman, *Appl. Spectrosc. Rev.*, 2007, **42**, 493-541.
- 10 [41]. I. Notingher, et al. *Anal. Chem.*, 2004, **76**, 3185-3193.
- 11 [42]. Z. Huang, et al. *Int. J. Cancer*, 2003, **107**, 1047-1052.
- 12 [43]. N. Stone, et al. *Faraday Discuss.*, 2004, **126**, 141-157.
- 13 [44]. T. Kohonen, *Biol. Cybern.*, 1982, **43**, 59-69.
- 14 [45]. T. Kohonen, et al. *Hels. Univ. Technol. S.*, 1996.
- 15 [46]. G. R. Lloyd, R. G. Brereton, and J. C. Duncan, *Analyst*, 2008, **133**, 1046-1059.
- 16 [47]. K. Wongravee, et al. *Anal. Chem.*, 2009, **82**, 628-638.
- 17 [48]. M. S. Bergholt, et al. *Technol. Cancer Res. T.*, 2011, **10**, 103-112.
- 18 [49]. S. K. Teh, et al. *Brit. J. Surg.*, 2010, **97**, 550-557.
- 19 [50]. I. Shima, et al. *Cancer*, 1992, **70**, 2747-2753.
- 20 [51]. Y. Adachi, et al. *Int. J. Oncol.*, 1998, **13**, 1031-1036.
- 21 [52]. S. Samantaray, et al. *J. Cancer Res. Clin.*, 2004, **130**, 37-44.
- 22 [53]. C. M. Alexander and Z. Werb, *Curr. Opin. Cell Biol.*, 1989, **1**, 974-982.
- 23 [54]. A. J. Docherty and G. Murphy, *Ann. Rheum. Dis.*, 1990, **49**, 469-479.
- 24 [55]. J. F. Woessner, *Faseb J.*, 1991, **5**, 2145-2154.
- 25 [56]. L. M. Matrisian, *Bioessays*, 1992, **14**, 455-463.
- 26 [57]. T. W. Stone and L. G. Darlington, *Nat. Rev. Drug Discov.*, 2002, **1**, 609-620.
- 27 [58]. K. Schröcksnadel, et al. *Clin. Chim. Acta*, 2006, **364**, 82-90.
- 28 [59]. N. K. Bassal, B. P. Hughes and M. Costabile, *J. Metabonomics Metabolites*, 2012,  
29 DOI:10.4172/jmbm.1000e107.
- 30 [60]. G. Zhang, et al. *Clin. Dev. Immunol.* 2011, DOI:10.1155/2011/384726.
- 31 [61]. X. Q. Liu and X. Wang, *Chin. Med. J.*, 2009, **122**, 3072-3077.
- 32 [62]. A. Huang, et al. *Brit. J. Cancer*, 2002, **86**, 1691-1696.
- 33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

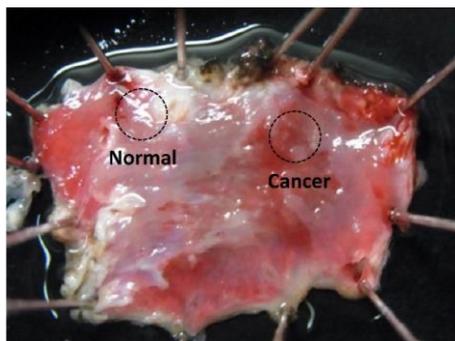


Figure 1: An image of esophageal tissue resected by endoscopic mucosal resection (EMR). It is a fresh *ex vivo* sample before the histopathological examinations. The tissue includes both normal and carcinoma areas.

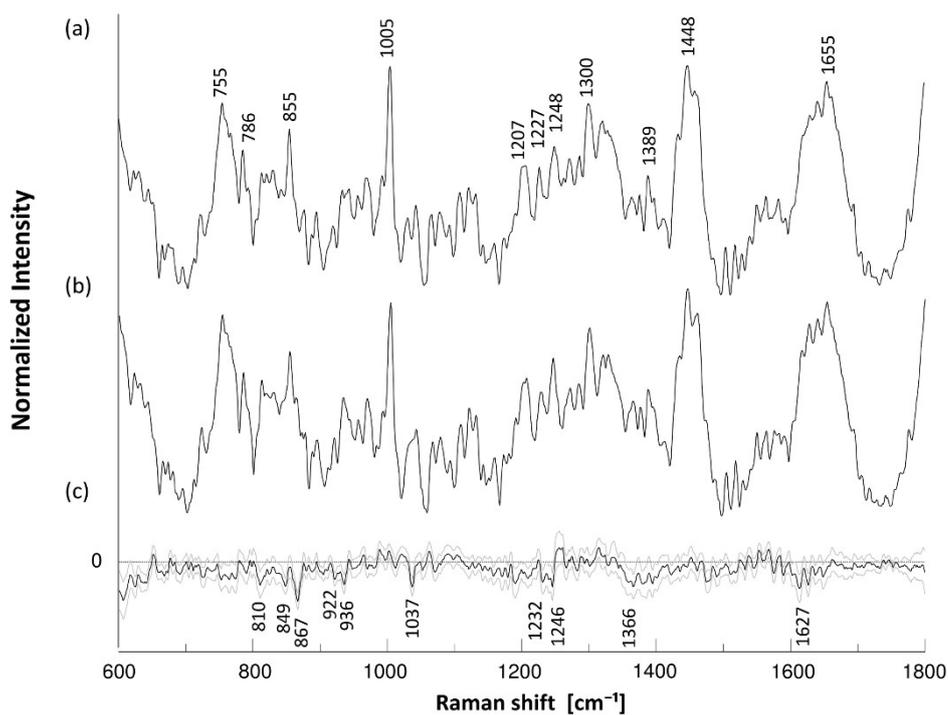


Figure 2: *Ex vivo* mean Raman spectra of (a) cancer (n=73: including stage I (n=42), stage 0 (n=25), and suspicious lesion (n=6)) and (b) normal (n=50) tissues, as well as (c) their subtraction spectra ((a)-(b)) with  $\pm 1$  standard errors.

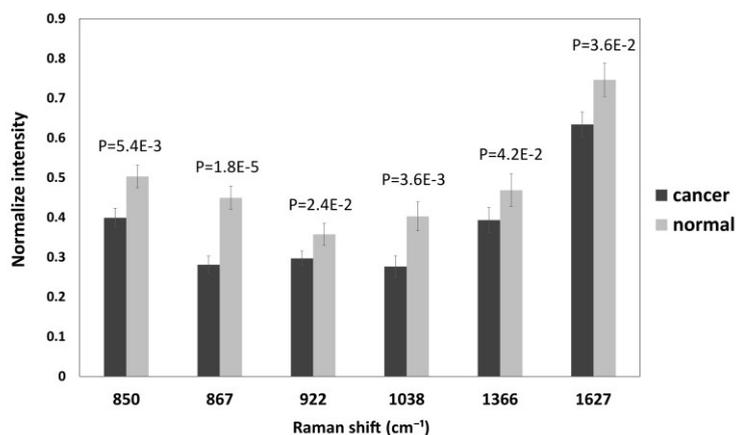


Figure 3: Six bands judged to have statistically significant differences between cancer and normal tissues by *t*-test.

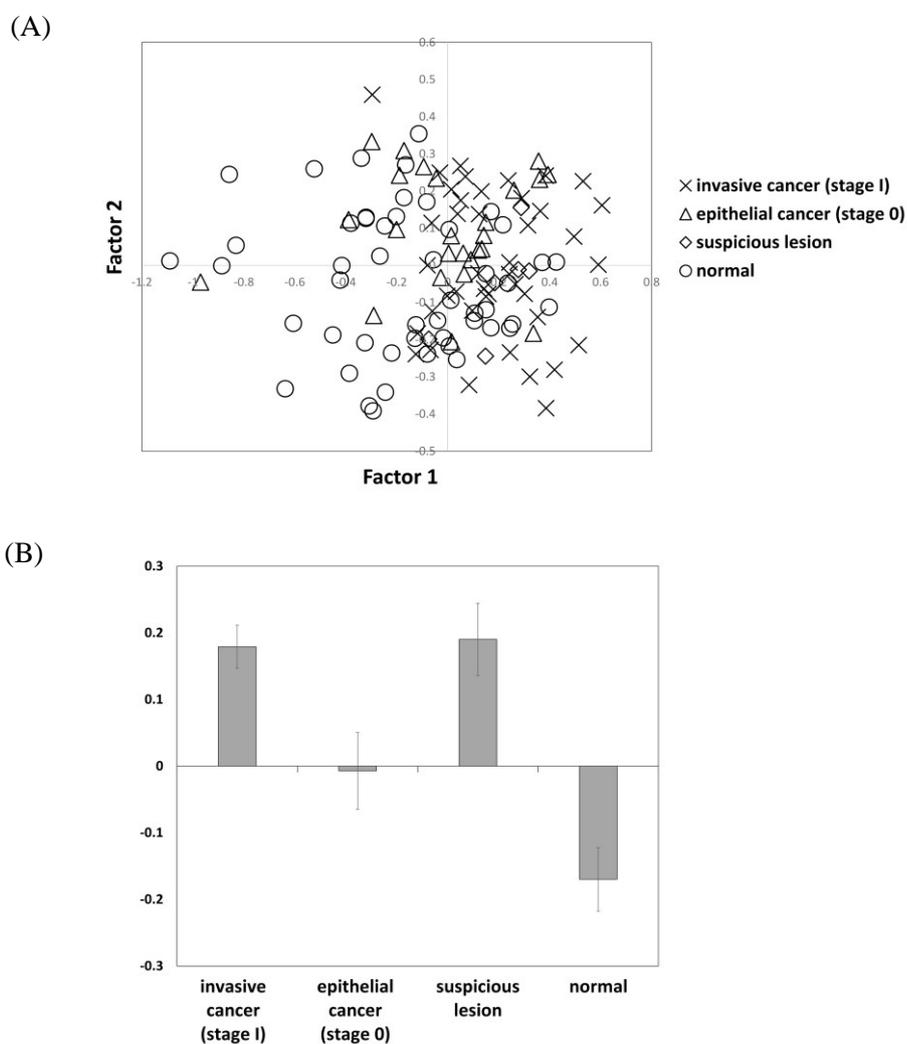


Figure 4: (A) PLS score plot obtained by using the selected wavenumbers from *t*-test. (B) The score average for each tissue type with standard error.

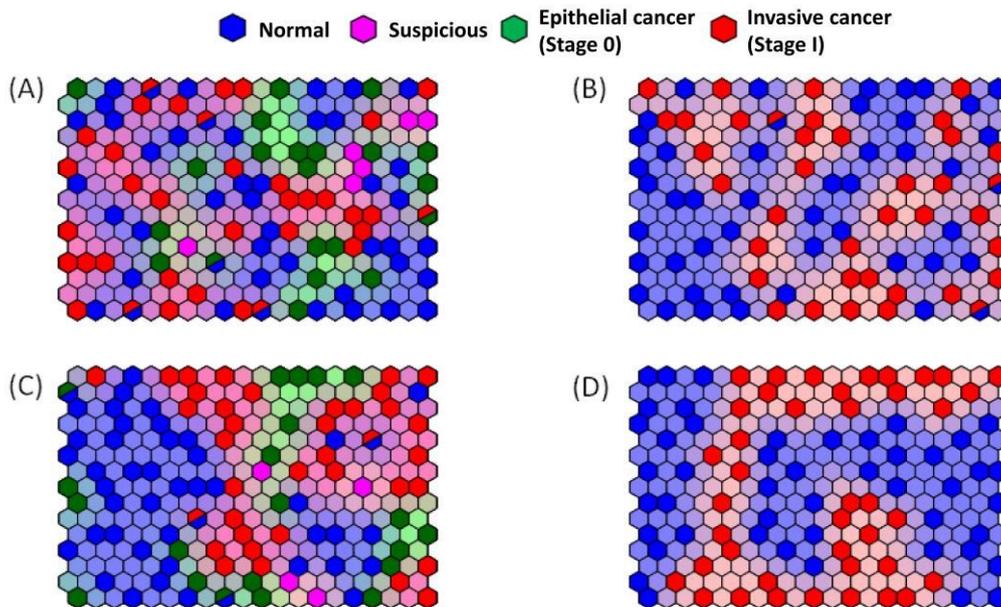


Figure 5: Self-Organization Maps (SOM) developed using (A)-(B) all wavenumbers and (C)-(D) only significant wavenumbers by  $t$ -test.

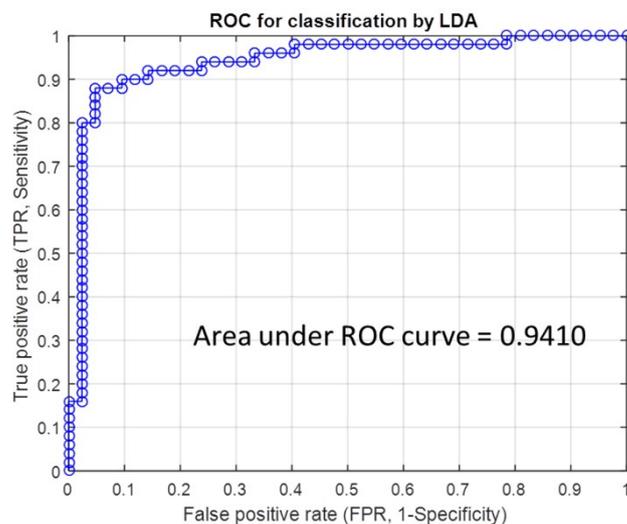


Figure 6: An ROC curve produced by plotting the proportion of true positive (TPs) against the proportion of false positive (FPs). The LDA diagnostic algorithm validated these two tissue types using the leave-one-data site-out, cross validation methodology.

		True class	
		Positive test + (Cancer)	Negative test - (Normal)
Prediction	Positive test + (Cancer)	True Positive (TP) 34	False Positive (FP) 3
	Negative test - (Normal)	False Negative (FN) 8	True Negative (TN) 47

Table 1: The validation result using LDA with LOOCV approach. The classifier were performed on the 92 sample data selected from normal and invasive cancer (stage I) tissues , those for 81 samples were correctly predicted, whereas only 11 samples were misjudged. Sensitivity ( $=TP/(TP+FN)$ ); 81.0 %, specificity ( $=TN/(TN+FP)$ ); 94.0%.