Gastric cancer differentiation using Fourier transform near-infrared spectroscopy with unsupervised pattern recognition

Wei-song Yi \textsuperscript{a,b}, Dian-sheng Cui \textsuperscript{c}, Zhi Li \textsuperscript{b}, Lan-lan Wu \textsuperscript{d}, Ai-guo Shen \textsuperscript{a,*}, Ji-ming Hu \textsuperscript{a,*}

\textsuperscript{a} College of Chemistry and Molecular Sciences, Wuhan University, Wuhan 430072, PR China
\textsuperscript{b} College of Science, Huazhong Agricultural University, Wuhan 430070, PR China
\textsuperscript{c} Gastric & Intestine Department, Hubei Cancer Hospital, Wuhan 430079, PR China
\textsuperscript{d} College of Technology, Huazhong Agricultural University, Wuhan 430070, PR China

\section*{Highlights}
- Major spectral differences were observed in three regions.
- Unsupervised pattern recognition techniques (PCA and CA) were used.
- The sensitivity, specificity and accuracy were 100%, 68.2% and 81.1%, respectively.

\section*{Article Info}

Article history:
Received 10 May 2012
Received in revised form 8 September 2012
Accepted 16 September 2012
Available online 26 September 2012

Keywords:
Gastric cancer
Differentiation
Fourier transform near-infrared spectroscopy
Unsupervised pattern recognition

\section*{Abstract}
The manuscript has investigated the application of near-infrared (NIR) spectroscopy for differentiation of gastric cancer. The 90 spectra from cancerous and normal tissues were collected from a total of 30 surgical specimens using Fourier transform near-infrared spectroscopy (FT-NIR) equipped with a fiber-optic probe. Major spectral differences were observed in the CH-stretching second overtone (9000–7000 cm\(^{-1}\)), CH-stretching first overtone (6000–5200 cm\(^{-1}\)), and CH-stretching combination (4500–4000 cm\(^{-1}\)) regions. By use of unsupervised pattern recognition, such as principal component analysis (PCA) and cluster analysis (CA), all spectra were classified into cancerous and normal tissue groups with accuracy up to 81.1%. The sensitivity and specificity was 100% and 68.2%, respectively. These present results indicate that CH-stretching first, combination band and second overtone regions can serve as diagnostic markers for gastric cancer.

\section*{Introduction}
Gastric cancer is rampant in many countries around the world [1]. By estimate, it is the fourth most common cancer worldwide [2]. Almost two-thirds of the cases occur in developing countries and 42% in China alone [3]. Survival for gastric cancer is moderately good only in Japan (52%), where mass screening by photofluoroscopy has been practiced since the 1960s. Survival is also relatively high in North America, possibly due to early diagnosis following a greater number of endoscopic examinations performed for gastric disorders [3]. The early detection and localization of abnormalities with appropriate curative treatment (e.g., endoscopic mucosal resection (EMR) and/or radiation therapy in combination with chemotherapy) are critical to decreasing the mortality rate [4]. However, the accurate identification and localization of early dysplasia and carcinoma in situ (CIS) and flat mucosal cancers in the gastric can be very challenging for clinicians using conventional white-light reflectance (WLR) endoscope [4].

While biopsy remains the “gold standard” approach for gastric cancer diagnosis, it is invasive and impractical as a routine screening tool for high-risk patients who may have multiple suspicious lesions. Furthermore, accurate identification of the tumor margins is of considerable clinical importance, in particular for the diffuse type of gastric cancer where the boundaries of the lesion can be indistinct and nests of cancerous cells may be found at a distance from the visible tumor margin. Therefore, the development of an advanced optical diagnostic technique that could identify and diagnose early cancerous lesions in vivo would be of significant clinical value during routine endoscopic inspections [5]. Thus, there are
needs to develop an accurate, fast, convenient and inexpensive method to screen gastric cancer. NIR spectroscopy, as a sensitive analytical technique with practical advantages [6,7], can record the response of chemical bonds in functional groups (e.g., O–H, C–H, and N–H bands) to the NIR spectrum, which is related to the primary structural components of organic molecules [8]. Therefore, any alteration in the composition of the tissues can be detected and used for diagnostic purposes.

Since cancer tissues differ from their normal parts in their composition, physiology, and biochemistry [9]. Several research groups have investigated the application of NIR spectroscopy for cancer differentiation in both animal models and human tissues. While NIR spectroscopy studies on animal models have mainly reported physiological aspects, studies on human tissues also focused on the compositional differences between tumor and normal tissue [10]. NIR spectroscopy studies on human tissues have been performed on breast [11,12], cervix [13], skin [14–16], prostate [17], brain [18], lung [19], head and neck [20], pancreas [21,22] and colorectal tissue [22,23]. In those papers they reported the use of NIR spectroscopy for diagnosis of among diverse cancers and draw attention to spectral differences revealed by use of different pattern-recognition methods. To our knowledge, there have been no reports on NIR spectroscopy differentiation of gastric cancer. In the present study, we therefore investigated the qualitative near-infrared spectral differences between gastric cancer and normal tissues in surgically resected gastric specimens using FT-NIR equipped with a fiber-optic probe to mimic in situ clinical measurements. In the present study, we therefore investigated the qualitative near-infrared spectral differences between gastric cancer and normal tissues in surgically resected gastric specimens using FT-NIR equipped with a fiber-optic probe to mimic in situ clinical measurements. In the present study, we therefore investigated the qualitative near-infrared spectral differences between gastric cancer and normal tissues in surgically resected gastric specimens using FT-NIR equipped with a fiber-optic probe to mimic in situ clinical measurements. In the present study, we therefore investigated the qualitative near-infrared spectral differences between gastric cancer and normal tissues in surgically resected gastric specimens using FT-NIR equipped with a fiber-optic probe to mimic in situ clinical measurements.

Materials and methods

Patient samples

Gastric cancer samples were collected immediately from 15 randomly selected patients who underwent partial gastric resection for tumor removal at Hubei Cancer Hospital. A total of 30 samples, two samples from each patient, were obtained. A cancer and normal (control) sample has been prepared for each patient. It was believed that the region with 5–10 cm distance from the cancer was to be healthy and that was confirmed by pathologic results. The gastric cancer samples were frozen in liquid nitrogen and then stored at −87 °C ultra-low temperature freezer until spectral analysis. A total of 15, eight males and seven females, the mean age of the patients was 56 years with the oldest 78 years and the youngest 29 years. The study was approved by the local ethics committee, and informed consent for use of samples was obtained from all patients.

NIR measurements

The spectrometer, Antaris II FT-NIR analyzer (Thermo Fisher Scientific, USA), furnished with an indium gallium arsenide (InGaAs) detector, was equipped with a fiber-optic module (SabIR Diffuse Reflectance Probe). The SabIR was a high-performance fiber-optic probe that enables remote non-destructive sampling. The head of the fiber optic probe had an outer diameter of 4.0 mm. The spot diameter of the probe was of 3.0 mm, and the approximate sampling area was 7.0 mm². This accessory was particular useful for raw material identification, quality measurements, and sample component analysis. The instrument was controlled with Result 3.0, a software package accompanying this analyzer.

The sample tissue was kept in a plastic plane, mucosa surface upwards and the fiber-optic probe was attached to a clamp. The tip of the probe was brought into contact with the mucosa surface of gastric tissue absolutely to mimic in situ clinical measurements. Each spectrum was the average spectrum of 32 scans [22,23]. The range of spectra was 10000–4000 cm⁻¹, and the raw spectra were measured in 4 cm⁻¹ interval, which resulted in 1500 variables. Three repeat measurements from each spot, every measurement was stored for further evaluation. A total of 90 spectra with log (1/R) as ordinate values (R is the sample diffuse reflectance) covering the two types of tissue were available for classification.

After measuring and marking the tissue surface, samples were fixed in 10% formalin solution and submitted to histopathological examination. Only spectra that were correctly acquired from the surfaces of tissues were used for data analysis after comparing with histopathological results [24].

Spectra preprocessing and data analysis

All spectra were acquired by use of the Result 3.0 software. Prior to the evaluation of the spectra using pattern recognition methods, the spectra preprocessing was performed by Savitzky–Golay smoothing filter, first derivative, and Norris derivative smoothing filter using the spectral analysis software OMNIC 8.0 (Thermo Fisher Scientific, USA). Spectral data were processed using the Scientific Graphing software Origin 8.0 (OriginLab, Northampton, MA, USA). After spectra preprocessing, spectral data were analyzed with PCA and CA, respectively. Concerning the unsupervised classification, samples are classified without a prior knowledge, except the spectra [25,26].

Principal component score plots can often reveal patterns or clustering within a data set. Therefore, PCA was done using the pattern recognition software Matlab R2010a (Math Work Inc., South Natick, MA, USA) to discriminate the cancer tissue spectra from normal tissue spectra.

In general, clustering is the partitioning of a data set into subsets (i.e. the clusters) so that the differences between the data within each cluster are minimized and the differences between clusters are maximized according to some defined distance measure [26]. CA was also done using the pattern recognition software Matlab R2010a to discriminate the cancer tissue spectra from normal tissue spectra.

Results and discussion

Mean spectra and band assignment

The absorption peaks of raw NIR spectra are broad and overlap, and it is impossible to make direct quantification analysis due to the high dimension and complexity of NIR spectral data [27]. Although NIR spectra cannot provided some significant differences of absorbance peaks like mid-infrared (MIR) spectroscopy, they still include much information about chemical composition of the tissue [8].

The differences in tissue composition between gastric cancer and normal tissue have been extensively analyzed using chemical, histochemical, biochemical and immunohistochemical methods [1]. These different tissue compositions will contribute to the spectral variance between cancer and normal tissue samples. The mean spectra of the cancer and normal tissue measurements from one sample are shown in Fig. 1. The variance is essentially constant across the spectral range used, that suggests the spectra are highly reproducible.

The near-infrared spectrum of a normal gastric tissue and the ordinate values (R) cover...
Spectra processing and data pre-processing

To reduce the differences due to tissue size and thickness, the spectra firstly were filtered by Savitzky–Golay smoothing. Then, spectra processing by first derivative were used to enlarge the spectral differences and remove any offset or linear baseline effects. Finally, the spectra were filtered by Norris derivative smoothing, as derivation amplifies the spectral noise. To emphasize the differences between the preprocessed cancer and normal spectra, the mean normal spectrum was subtracted from the mean cancer spectrum.

The preprocessed mean spectrum of cancer and normal tissue and their differences are shown in Fig. 2. Distinct differences in the spectra were observed at three intervals, the CH-stretching second overtone (9000–7000 cm\(^{-1}\)), CH-stretching first overtone (6000–5200 cm\(^{-1}\)), and CH-stretching combination (4500–4000 cm\(^{-1}\)) regions, illustrate the major abnormalities. The similar discrepancy of near-infrared spectrum for the diagnosis of colorectal cancer has been reported by Kondepati et al. [23]. The similarity has indicated the coherence of carcinoma differentiation by NIR.

Data evaluation with unsupervised pattern recognition

Any alteration in the chemical composition of the tissues can thus be detected by NIR spectroscopy and evaluated with the help of effective unsupervised pattern recognition techniques [23].

Principal component analysis

To reduce the dimension of the spectral data, PCA can be employed to extract a set of orthogonal principal components comprising loadings and scores that account for the maximum variance in spectral datasets [28]. Principal component score plots can often reveal patterns or clustering within a data set. Fig. 3 shows the first versus second principal component scores obtained from all spectra of cancer and normal tissue for the evaluations based on the CH-stretching second overtone, CH-stretching first, and CH-stretching combination regions. Clear clustering of the data for the two distinct groups can be observed. This provides strong evidence that significant molecular differences between normal and cancer tissue can be revealed with NIR spectroscopy.

It is also noticeable that the compositional variability in these spectral regions is more for the cancer tissues. The greater scatter in the cancerous tissue spectra in the CH-stretching second overtone, CH-stretching first, and CH-stretching combination regions reflects the larger compositional variability in the cancerous tissues. The relatively high variation between the cancer samples could be due to different tumor stages of the tissues and differences in the tissue thickness of the samples affecting the shape of the diffuse reflectance spectra due to different photon tissue
two regions, which can be due to operational errors [29]. NIR spectra. These spectral outliers were not common for the outliers. Only two spectra were outliers within the pane for all regions can serve as diagnostic marker for gastric cancer.

These findings suggest that NIR spectroscopy is a sensitive tool for the detection of gastric cancer.

Conclusions

The differences from chemical characteristics of cancerous and normal tissues reasonably differentiated in the NIR spectroscopy. Major spectral differences were observed in the CH-stretching second overtone, CH-stretching first, and CH-stretching combination regions. Therefore, NIR spectra can exhibit the neat cluster trend of gastric specimens according to normal and cancerous status by means of principal component analysis. Moreover, using unsupervised pattern recognition the spectra were classified into cancer and normal tissue groups with an inherent high accuracy and specificity, without the need for costly and laborious chemical analysis. These present results suggest that FT-NIR spectroscopy is a simple, feasible and sensitive method for screening gastric cancer.

Acknowledgements

This work was supported by National Natural and Science Foundation of China (Nos. 21175101, 20927003, 90913013, and 20805034) and Foundation of Chemo/Bio sensing and Chemometrics, Hunan University. The author (W.S. Yi) gratefully acknowledges the support of the Fundamental Research Funds for the Central Universities (No. 2011JC017).

References
