

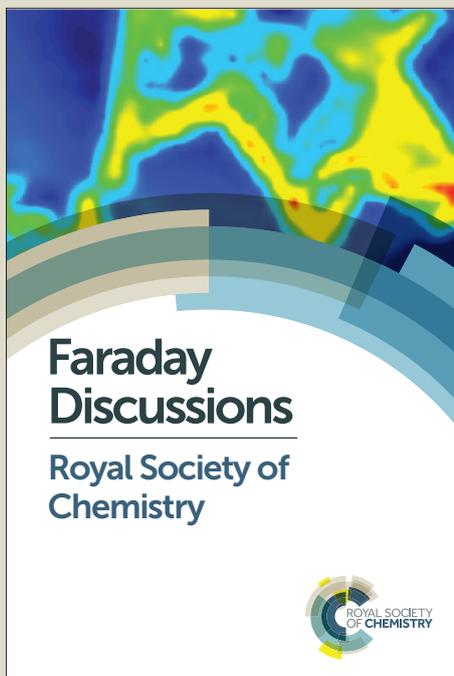
Faraday Discussions

Accepted Manuscript



This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting. All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: <http://rsc.li/fd-upcoming-meetings>



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.

Fiber-optic Raman spectroscopy for *in vivo* diagnosis of gastric dysplasia

Jianfeng Wang^{1*}, Kan Lin^{1*}, Wei Zheng^{1*}, Khek Yu Ho², Ming Teh³,
Khay Guan Yeoh², and Zhiwei Huang^{1**}

¹Optical Bioimaging Laboratory, Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, Singapore 117576

²Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 119260

³Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 119074

**Wang, Lin and Zheng contributed equally*

Running Title: Raman spectroscopy detects gastric dysplasia *in vivo*

Key words: Fiber-optic Raman spectroscopy · *In vivo* diagnosis · Gastric dysplasia .

**Corresponding author and address for reprints:

Dr. Zhiwei Huang

Optical Bioimaging Laboratory,

Department of Biomedical Engineering

Faculty of Engineering, National University of Singapore

9 Engineering Drive 1, Singapore 117576

Tel: +65 6516 8856, Fax: +65 6872 3069, E-mail: biehzw@nus.edu.sg

Abstract

This study aims to assess clinical utility of a rapid fiber-optic Raman spectroscopy technique developed for enhancing *in vivo* diagnosis of gastric precancer during endoscopic examination. We have developed a real-time fiber-optic Raman spectroscopy system capable of simultaneously acquiring both fingerprint (FP) (i.e., 800–1800 cm^{-1}) and high-wavenumber (HW) (i.e., 2800–3600 cm^{-1}) Raman spectra from gastric tissue *in vivo* at endoscopy. A total of 5792 high-quality *in vivo* FP/HW Raman spectra (normal (n=5160); dysplasia (n=155), and adenocarcinoma (n=477)) were acquired in real-time from 441 tissue sites (normal (n=396); dysplasia (n=11), and adenocarcinoma (n=34)) of 191 gastric patients (normal (n=172); dysplasia (n=6), and adenocarcinoma (n=13)) undergoing routine endoscopic examinations. Partial least squares discriminant analysis (PLS-DA) together with leave-one patient-out, cross validation (LOPCV) were implemented to develop robust spectral diagnostic models. FP/HW Raman spectra differ significantly between normal, dysplasia and adenocarcinoma of the stomach that can be attributed to changes in proteins, lipids, nucleic acids, and the bound water content. PLS-DA and LOPCV shows that the fiber-optic FP/HW Raman spectroscopy provides a diagnostic sensitivities of 96.0%, 81.8% and 88.2%, and specificities of 86.7%, 95.3% and 95.6%, respectively, for the classification of normal, dysplastic and cancerous gastric tissue, superior to either FP or HW Raman technique alone. Further dichotomous PLS-DA analysis yields a sensitivity of 90.9% (10/11) and specificity of 95.9% (380/396) for the detection of gastric dysplasia using FP/HW Raman spectroscopy, substantiating its clinical advantages over white light reflectance endoscopy (sensitivity: 90.9% (10/11), and specificity: 51.0% (202/396)). This work demonstrates that the fiber-optic FP/HW Raman spectroscopy technique has great promise for enhancing *in vivo* diagnosis of gastric precancer during routine endoscopic examination.

Introduction

Gastric cancer is the fifth most commonly diagnosed malignancies and also the third leading cause of cancer death worldwide, with an estimated 951,600 new cases and 723,100 deaths occurred in 2012, accounting for 6.8%, and 8.8%, respectively of all the estimated new cancer cases (14,090,100) and deaths (8,201,600) ¹. The incidence rates of gastric cancer are highest in Eastern Asia (e.g., Japan, China, South Korea, Mongolia) with an incidence rate of 35.4% in males and 13.8% in females ¹. Early detection and localization with immediate removal and treatment (e.g., endoscopic submucosal dissection (ESD)) of premalignant lesions (e.g., high-grade dysplasia [HGD]) is crucial to improving survival of gastric patients who underwent endoscopic screening of the stomach ²⁻⁵. Current routine diagnostics of gastric dysplasia is clinically challenging, as conventional white-light reflectance (WLR) endoscopy used relies heavily on visual identification of gross morphologic changes. While in 60% of the gastric patients, gastric dysplasia was shown as flat mucosa in the absence of specific endoscopic alterations, resulting in a limited diagnostic accuracy by using WLR endoscopy ³⁻⁵. Well-performed but randomized endoscopic biopsies followed by conventional hematoxylin and eosin (H&E) staining remains the current gold standard for gastric cancer screening and diagnosis ³⁻⁵. However, random biopsies associated with sampling errors make it labor-intensive and difficult for screening of high-risk gastric patients with multiple suspicious lesions ³⁻⁵. Further, the diagnosis is also compromised with intra- and/or inter- operators' variations ³⁻⁵. There is profound clinical need to develop more advanced diagnostic techniques, providing rapid,

objective and enhanced diagnosis of gastric precancer (dysplasia) with molecular specificity.

Optical spectroscopic techniques (e.g., diffuse reflectance, autofluorescence, and Raman spectroscopy) cast new light on either uncovering tissue optical properties (e.g., absorption and scattering coefficients), endogenous fluorophore distribution or the constituent-specific (i.e., DNA, proteins, lipids, and water etc) biochemical/biomolecular information which offer objective insights, and therefore have attracted a great deal of interest for precancer and cancer diagnosis in internal organs (e.g., stomach, esophagus, colon, bladder, and lung, etc) ⁶⁻²⁸. Raman spectroscopy is a label-free optical vibrational technique capable of probing biochemical/biomolecular changes of tissue associated with neoplastic transformation ^{6-27, 29}. Currently, the exploration of biomedical Raman spectroscopy for gastric cancer and precancer diagnosis is mostly centered on the fingerprint (FP) range (i.e., 800-1800 cm^{-1}) which uncovers a wealth of biochemical information (i.e., proteins, lipids and nucleic acid) in cells and tissue ^{8-20, 27, 28}. However, the FP Raman spectroscopy could suffer from relatively weak Raman intensity but overwhelmed tissue autofluorescence background in certain organ sites (e.g., stomach, lung, liver) ^{12, 20}. High-wavenumber (HW) (e.g., 2800–3600 cm^{-1}) Raman spectroscopy, on the other hand, can provide complementary biochemical information that is not achievable with the FP Raman spectroscopy (e.g., tissue water content), while has stronger tissue Raman signals with less autofluorescence interference ²¹⁻²⁴. Several rationales exist by combining the FP and HW Raman spectroscopy for gastric precancer and early cancer

diagnosis at endoscopy: For tissues (e.g., stomach) exhibiting intense autofluorescence overwhelming the tissue FP Raman signals, the HW Raman spectroscopy could still contain reasonably intense tissue HW Raman peaks associated with gastric tissue neoplasia; Further, the FP and HW Raman spectroscopy are complementary, and the integrated FP/HW Raman spectroscopy could have advantage for improving tissue characterization and diagnosis^{25,26}. Very recently, we have successfully developed an endoscope-based fiber-optic Raman spectroscopy system capable of simultaneously acquiring both the fingerprint (FP) (i.e., 800–1800 cm^{-1}) and high-wavenumber (HW) (i.e., 2800–3600 cm^{-1}) tissue Raman spectra *in vivo* at endoscopy^{25, 26}. In this work, we investigate the clinical utility of the integrated FP/HW Raman spectroscopy technique developed to enhance real-time *in vivo* diagnosis of gastric dysplasia during clinical endoscopic examinations. Partial least squares - discriminant analysis (PLS-DA) and leave-one patient-out, cross validation (LOPCV) are implemented to develop robust spectral diagnostic algorithms for the differentiation between dysplasia and normal gastric tissue. The diagnostic performances of the FP/HW Raman spectroscopy technique are also compared with the routine WLR endoscopic examination.

Materials and Methods

Raman Instrumentation and Data Processing. We have developed a rapid fiber-optic FP/HW Raman endoscopic system^{25,26} that consists of a 785 nm diode laser (maximum output: 300 mW, B&W TEK Inc.), a customized beveled fiber-optic Raman probe (1.8 mm in outer diameter)²⁷, and a high-throughput reflective imaging spectrograph (Acton LS-785 f/2, Princeton Instruments Inc.) equipped with a customized high-diffraction efficiency (>85% within in range 800-3600 cm⁻¹) gold-coated grating (830 gr/mm) and a thermo electric-cooled, NIR-optimized deep-depletion charge-coupled device (CCD) camera (PIXIS 400BR eXcelon, Princeton Instruments Inc.) for tissue Raman signals. The beveled fiber-optic Raman endoscopic probe (1.8 mm in outer diameter) compatible with most of medical endoscopes was specially designed and fabricated for optimizing epithelial tissue Raman measurements²⁷: i.e., ~85% of the FP/HW Raman signal collected by the beveled Raman endoscopic probe is arising from the top ~200 μm epithelium layer; whereas ~15% of the Raman collection is from below ~(200-800) μm of the gastric tissue, facilitating the *in vivo* diagnosis of gastric precancer²⁷. The 785-nm laser power incident on the tissue surface is ~12 mW (equivalent to ~1.5 W/cm² within the spot size of ~500 μm) permissible by the American National Standards Institute (ANSI) standard³⁰. To further evaluate the photothermal effect of the laser power used, thermal modeling was conducted on the stomach³¹⁻³⁴, showing a non-significant temperature rise (<0.07 °C) without causing photothermal damage to the gastric tissue within 1 minute of laser irradiation time³⁵. The system was wavelength-calibrated

using a mercury/argon calibration lamp (HG-1 and AR-1, Ocean Optics Inc., Dunedin, FL). The spectral response correction for the wavelength-calibrated system was conducted using a standard lamp (RS-10, EG&G Gamma Scientific, San Diego, CA)²⁸.

The measured gastric tissue spectra represented a combination of weak tissue Raman scattering, intense AF and the noise. Therefore, the spectra were preprocessed by a third-order Savitzky-Golay smoothing filter (a window width of 3 pixels) to reduce the noise. A fifth-order polynomial was found to be optimal for fitting the AF background in the noise-smoothed spectrum over the range of 800-1800 cm^{-1} , and this polynomial was then subtracted from the measured FP spectrum to yield the FP tissue Raman spectrum alone. While in the HW range (2800-3600 cm^{-1}), the first-order polynomial fit was used for removing the HW AF background. The tissue FP/HW Raman spectrum was finally normalized to the integrated area of the entire FP and HW range to compare the differences in spectral shapes and relative Raman band intensities between dysplastic and normal gastric tissue. One notes that there are no Raman peaks observed in the wavenumber range of 1800-2800 cm^{-1} , and this silent range is therefore excluded for data analysis. All the preprocessing is completed within 30 ms, and the processed Raman spectra and diagnostic outcomes can be displayed on the computer screen in real-time during clinical endoscopic examinations²⁸.

Clinical trial protocol. The ethical protocol of the present study was approved by the Institutional Review Board (IRB) of the National Healthcare Group (NHG) of Singapore. Prior to Raman measurements, all patients (21 to 80 years old) signed an informed consent permitting the *in vivo* Raman spectroscopic measurements during gastric endoscope examinations. The trial has been registered (registration number: ISRCTN15587241) on June 23, 2015 into ISRCTN registry (<http://www.isrctn.com/>), which is a publicly accessible primary register that participates in the World Health Organization (WHO) International Clinical Trial Registry Platform. Also, the trial was performed in accordance with International Conference on Harmonization (ICH) for Good Clinical Practice (GCP) guidelines, Declaration of Helsinki (2000). A total of 191 gastric patients (115 men and 76 women with mean ages of 55; normal (n=172), dysplasia (n=6), and adenocarcinoma (n=13)) were enrolled in the gastrointestinal Raman endoscopic examinations. The Raman probe passes down to the instrument channel of medical endoscopes (GIF FQ260Z; Olympus Medical Systems, Tokyo, Japan) under the guidance of WLR endoscopic imaging¹³. The Raman probe tip was visible approximately 0.5 cm in front of the endoscope camera. During endoscopic examination of suspicious lesions, the Raman probe was placed in gentle contact with the gastric mucosa surface, and the positioning of the Raman probe against the tissue sites was verified on the computer monitor by the endoscopists in-charge at endoscopy. Multiple spectra (~5-15) for each tissue site were measured with scanning times of 0.1 to 0.5 sec, which permits a rapid survey of large tissue areas. Raman spectra acquired in non-contact with gastric tissues were

automatically discarded online using Raman diagnostic software developed by our group²⁸. Immediately after the tissue Raman acquisitions, each tissue site measured was biopsied and sent for histopathological examination by three senior gastrointestinal pathologists who were blinded to the Raman scans. The consented histopathology assessments on the biopsied tissues serve as the gold standard to determine the diagnostic ability of the FP/HW fiber-optic Raman technique for identifying gastric dysplasia from normal gastric tissue.

Statistical Analysis. Cohen's κ statistics were calculated to examine the agreement among the three pathologists³⁶. One-way analysis of variance (ANOVA) with Fisher post hoc least significant difference (LSD) test was used to evaluate the Raman spectral differences among the gastric tissue subtypes (i.e., normal, dysplasia and adenocarcinoma)²⁸. Partial least squares (PLS) - discriminant analysis (DA) was applied for developing spectral diagnosis models²⁶. Leave-one-patient out, cross-validation was further used to assess and optimize the PLS-DA model complexity, while reducing the risk of over-fitting. The overall discriminatory accuracy of the FP, HW, and the integrated FP/HW diagnostic models were evaluated through the use of the receiver operating characteristic (ROC) analysis³⁷. DeLong test was implemented on the ROC curves of the three (i.e., FP, HW, and the integrated FP/HW) diagnostic models to evaluate whether their areas under the curve (AUC) are statistically significant³⁸. The majority voting strategy was applied for final classification of each tissue site measured. The above multivariate statistical analysis

was performed using in-house written scripts in the Matlab programming environment (Mathworks. Inc., Natick, MA).

Results

A total of 5792 FP/HW gastric tissue Raman spectra (normal (n=5160), HGD (high-grade dysplasia, n=155), and adenocarcinoma (n=477)) were successfully acquired in real-time *in vivo* from 441 tissue sites (normal (n=396), HGD (n=11), and adenocarcinoma (n=34)) of the 191 gastric patients recruited (normal (n=172), HGD (n=6), and adenocarcinoma (n=13)). Fig. 1a shows the mean *in vivo* FP/HW Raman spectra \pm 1 standard deviation (SD) (shaded area) of normal, dysplasia and adenocarcinoma gastric tissues. Prominent gastric tissue Raman peaks with tentative assignments can be observed in the FP range⁸⁻²⁰, i.e., 853 (ν (C-C) proteins), 1004 (ν_s (C-C) ring breathing of phenylalanine), 1078 (ν (C-C) of lipids), 1265 (amide III ν (C-N) and δ (N-H) of proteins), 1302 (CH₂ twisting and wagging of lipids), 1335 (CH₃CH₂ twisting of proteins and nucleic acids), 1445 (δ (CH₂) deformation of proteins and lipids), 1618 (ν (C=C) of porphyrins), 1655 (amide I ν (C=O) of proteins) and 1745 cm⁻¹ (ν (C=O) of phospholipids). Intense Raman peaks with assignments are also observed in the HW range²¹⁻²⁶, i.e., 2850 and 2885 cm⁻¹ (symmetric and asymmetric CH₂ stretching of lipids), 2940 cm⁻¹ (CH₃ stretching of proteins), 3009 cm⁻¹ (asymmetric =CH stretching of lipid), ~3300 cm⁻¹ (amide A (NH stretching of proteins)) and the broad Raman band of water (OH stretching vibrations peaking at ~3250 and ~3400 cm⁻¹) that are related to the local conformation and interactions of

OH-bonds in the intracellular and extracellular space of gastric tissue. One notes that the intense broad Raman band of water above 3000 cm^{-1} has also been observed in other soft tissues (e.g., brain and oral cavity)^{39, 40}. Fig. 1b shows the difference Raman spectra (i.e., normal - dysplasia, normal - adenocarcinoma, dysplasia - adenocarcinoma) \pm 1SD (shaded area), reflecting the different Raman-active component contribution for different gastric lesions. The significant difference ($p=1.2\text{E-}9$, one-way analysis of variance (ANOVA) with Fisher post hoc least significant difference (LSD)) in Raman spectra of normal and neoplastic gastric tissue discerned (Fig. 1b) demonstrates the potential of the integrated FP/HW fiber-optic Raman spectroscopy for real-time *in vivo* diagnosis of gastric dysplasia and adenocarcinoma.

To elucidate the diagnostically important Raman-active components, Fig. 2a shows a logarithmic plot of the calculated p -values (one-way analysis of variance (ANOVA) with Fisher post hoc least significant difference (LSD) test at the 0.05 level.) for each of the Raman intensities in the entire spectral range (i.e., $800\text{--}1800\text{ cm}^{-1}$ and $2800\text{--}3600\text{ cm}^{-1}$ with a set of 779 intensities). In particular, nine spectral sub-regions with statically significant difference ($p<1\text{E-}9$) among normal gastric mucosa, dysplasia and gastric adenocarcinoma were found: i.e., $830\text{--}940\text{ cm}^{-1}$, $1000\text{--}1100\text{ cm}^{-1}$, $1176\text{--}1220\text{ cm}^{-1}$, $1256\text{--}1337\text{ cm}^{-1}$, $1385\text{--}1410\text{ cm}^{-1}$, $1440\text{--}1480\text{ cm}^{-1}$, $1600\text{--}1670\text{ cm}^{-1}$, and $2860\text{--}3000\text{ cm}^{-1}$, related to proteins, lipids and nucleic acids. Significant spectral differences were also observed in bound water in the ranges of $3160\text{--}3260\text{ cm}^{-1}$. Fig. 2b displays a histogram of the most statistically different Raman

peak intensities (mean \pm 1SD) from both FP and HW ranges at 853 cm^{-1} , 1004 cm^{-1} , 1078 cm^{-1} , 1210 cm^{-1} , 1265 cm^{-1} , 1302 cm^{-1} , 1335 cm^{-1} , 1445 cm^{-1} , 1618 cm^{-1} , 1655 cm^{-1} , 2850 cm^{-1} , 2885 cm^{-1} , 2940 cm^{-1} , 3250 cm^{-1} , and 3300 cm^{-1} , confirming the capability of the rapid FP/HW fiber-optic Raman spectroscopy to reveal the biochemical/biomolecular constituents (e.g., proteins, lipids, DNA and water *et al*) associated with gastric carcinogenesis *in vivo*. Besides, we have also correlated the *in vivo* Raman spectra with representative histopathology slides of the three different pathology categories (Fig. 3) including (a) normal gastric mucosa; (b) high-grade dysplasia showing abrupt transition of dysplastic epithelium, and (c) gastric adenocarcinoma. While the histopathology identifies prominent cellular and architectural anomalies associated with gastric carcinogenesis (Fig. 3), FP/HW Raman spectra uncovers the molecular specificity inherent with the gastric dysplasia and adenocarcinoma (Fig. 1).

To further develop robust multivariate spectral diagnostic algorithms and concurrently gain deeper insights into the underlying molecular biomarkers accompanied with the gastric neoplasia development, PLS-DA together with LOPCV was applied on the *in vivo* tissue Raman spectra acquired (Fig. 1). Essentially, the PLS-DA approach assigns orthogonal latent variables (LVs) on the basis of the spectral variance that correlate with the gold standard of histopathology (i.e., normal, dysplasia, and adenocarcinoma. Fig. 3). The molecular features associated with gastric carcinogenesis can be further extracted and visualized through loadings (Fig. 4). Loading on LV1 largely captured specific FP Raman peaks arising from saturated

and unsaturated lipid moieties (i.e., 1078 cm^{-1} ($\nu(\text{C-C})$), 1302 cm^{-1} (CH_2 twisting and wagging), 1445 cm^{-1} ($\delta(\text{CH}_2)$ deformation), 1655 cm^{-1} ($\nu(\text{C=C})$), 1745 cm^{-1} ($\nu(\text{C=O})$), and HW Raman peaks at 2850 and 2885 cm^{-1} (symmetric and asymmetric CH_2 stretching), 3009 cm^{-1} (asymmetric $=\text{CH}$ stretching of lipids) as well as the symmetric and asymmetric OH vibrations in the HW range ²¹⁻²⁶. On the other hand, loading on LV2 generally captured Raman peaks assigned to proteins and nucleic acids (i.e., 1004 cm^{-1} ($\nu_s(\text{C-C})$), 1265 cm^{-1} (amide III $\nu(\text{C-N})$ and $\delta(\text{N-H})$), 1335 cm^{-1} (nucleic acids), 1445 cm^{-1} ($\delta(\text{CH}_2)$ deformation), 1655 cm^{-1} (amide I $\nu(\text{C=O})$), 2940 cm^{-1} (CH_3 stretching), and ~ 3329 cm^{-1} (Amide A (N-H stretching)).

Fig. 5 shows the 2-dimensional ternary plot of the posterior probabilities of each Raman prediction using PLS-DA-LOPCV and majority voting for (a) FP, (b) HW, and (c) integrated FP/HW Raman spectroscopy, respectively. A good consistency for the gastric tissue groupings (i.e., normal, dysplasia and adenocarcinoma) was obtained among the three independent pathologists as indicated by the Cohen's kappa of 0.89 ³⁶. The prediction results were also summarized in Tables 1-3. We found that normal mucosa could be detected with accuracy of 95.0% (sensitivity: 96.0%, specificity: 86.7%), dysplastic tissues was identified with accuracy of 99.5% (sensitivity: 81.8%, specificity: 95.3%), whereas adenocarcinoma was identified with accuracy of 95.5% (sensitivity: 88.2%, specificity: 95.6%) by using integrated FP/HW Raman spectroscopy (Table 1), superior to using either FP (Table 2) or HW (Table 3) Raman technique alone.

We further conducted a dichotomous PLS-DA analysis to examine whether the

FP and HW Raman modalities were complementary for distinguishing the gastric dysplasia from normal gastric tissue. The integrated FP/HW Raman spectroscopy yielded a diagnostic accuracy of 95.7% [sensitivity of 88.4% (137/155) and specificity of 95.9% (4950/5160)], higher than those using either FP (accuracy of 92.8%; sensitivity of 79.3% (123/155), and specificity of 93.2% (4809/5160)) or HW (accuracy of 93.3%; sensitivity of 63.2% (98/155), and specificity of 94.2% (4861/5160)) Raman technique alone. The receiver operating characteristic (ROC) curves for the differentiation between dysplastic and normal gastric tissue are also generated (Fig. 6), with the integration areas under the ROC curves of being 0.935, 0.850 and 0.982 for the FP, HW, and the integrated FP/HW Raman techniques, respectively. Further DeLong test on the ROC-AUCs shows that the AUC of integrated FP/HW Raman technique is significantly higher than that of either FP ($p = 1.4E-3$, DeLong test) or HW ($p = 1.2E-6$, DeLong test) Raman technique³⁸, illustrating the diagnostic advantages of the integrated FP/HW Raman technique over either the FP or HW Raman technique. Overall, the above results reaffirm that integrated FP/HW Raman technique provides the best diagnostic performances for the *in vivo* detection of gastric dysplasia as compared to FP or HW Raman spectroscopy technique.

Discussion

We have applied an integrated FP/HW fiber-optic Raman spectroscopy system to acquire high-quality *in vivo* Raman spectra of gastric tissue in real time during clinical

endoscopic examinations. One main advantage of this system over prior Raman techniques is its ability to acquire both the FP Raman spectra and its complementary HW Raman spectra simultaneously at endoscopy. While histopathology (Fig. 3) identifies the abrupt transition of the dysplastic epithelium, the integrated FP/HW Raman spectroscopy reveals the biochemical and biomolecular changes occurring in the epithelium. Significant Raman spectral differences were observed between different pathological subtypes (Fig. 1b, and Fig. 2), confirming the diagnostic potential of the integrated FP/HW Raman spectroscopy for *in vivo* diagnosis of preneoplastic and neoplastic gastric lesions. FP Raman spectra reveal significant biochemical/biomolecular differences associated with gastric carcinogenesis. For instance, the increased thickness of the epithelial layer of gastric dysplasia attenuates the laser power delivered and also obscures the Raman signals from deeper subcutaneous fat layers⁴¹, leading to a significantly reduced Raman peaking at 1078 cm^{-1} attributed to phospholipids. Further, a decreased percentage of phenylalanine is indicated by the Raman peak at 1004 cm^{-1} . This finding is consistent with the previous study which established a link between essential amino acids deficiency (e.g. phenylalanine) and carcinogenesis⁴². We also found that gastric dysplasia is characterized by increased Raman peaks assigned to proteins (e.g., 1265, 1445, and 1655 cm^{-1}). This observation is in agreement with cytometric studies that dysplastic gastric tissues are accompanied with enhanced metabolic activities and increased hyperchromatism⁴³⁻⁴⁵. Specifically, dysplastic hyperchromatism induced an elevated concentration of histones, the main protein component that makes up the chromatin

for dysplasia tissue as indicated by the stronger Raman intensity at 1445 cm^{-1} .^{4,46} The increased histone concentration is also supported by the relative increased Raman-active components of amide III band (1265 cm^{-1}) and amide I band (1655 cm^{-1}) which are indicative of α -helix structures⁴⁷ (Fig. 1). Further comparison of the difference Raman spectra (Fig. 1b) *in vivo* brings the aggregated β -pleated sheet to light, which are indicative of enhanced biochemical interactions between the proteins and the cell microenvironment related to the increase of mitotic activity, one of the cellular alteration characteristics of gastric dysplasia⁴⁸. Besides, the Raman band at 1335 cm^{-1} (CH_3CH_2 twisting of proteins and nucleic acids) is higher for gastric dysplasia, signifying a dramatically increased DNA contribution to the development of gastric dysplasia. This is probably due to the fact that the dysplastic epithelium contains enlarged, hyperchromatic and crowded nuclei⁴⁵. Besides, the increased DNA observed are also consistent with DNA changes associated with tissue precancer and cancer transformation interrogated by using Raman spectroscopy technique^{8-20, 25, 26}. While the FP Raman spectra show specific but relatively weak Raman peaks that are predominantly related to backbone structures of proteins, lipids and DNA⁸⁻²⁰. The HW Raman spectra, on the other hand, offer exclusive new insights into the CH_2 and CH_3 stretching information and structure as well as the interactions of intra- and inter-cellular water *in situ*. The overall intensities of Raman signals of gastric dysplasia in the $2800\text{-}3020\text{ cm}^{-1}$ range involving proteins (e.g., 2940 cm^{-1}) and lipids (e.g., 2850 and 2885 cm^{-1}) are significantly higher than those in normal gastric tissue. The proteins to lipids Raman peak ratio (i.e., $I_{2940}/I_{2850} \pm \text{SD}$) reveals a significant

increase ($p < 0.001$) from normal (1.996 ± 0.009) to dysplastic (2.434 ± 0.001) gastric tissue, a well-known indicator of increased proteins components while simultaneously reduced lipids contents associated with the gastric carcinogenesis^{4, 5, 11, 19}. Moreover, the HW Raman spectrum resolves the changes of the subtle =CH stretch mode at 3009 cm^{-1} of gastric dysplasia that provides unequivocal evidence as a marker for saturated fatty acids (Fig. 2)⁴⁹. Such insights are crucial to deciphering the increasingly appreciated role of lipid metabolism in cancerous tissues⁴⁹. With the integrated FP/HW fiber-optic Raman spectroscopy, it is now possible to further strengthen the link between lipid metabolism and carcinogenic onset and progression *in situ*. In particular, prominently increased water content was also observed for the dysplastic and cancerous stomach as indicated by the significantly increased Raman peaking at 3250 cm^{-1} (Fig. 2b), reflecting the re-arrangement of hydrogen-bonded networks in the epithelium layer caused by local interaction with macromolecules such as collagen⁵⁰. Hence, the integrated FP/HW Raman endoscopy technique is not only capable of harvesting a wealth of information on proteins, lipids, and nucleic acids, but also provides information on the bound water content in gastric tissue for improving early diagnosis of gastric precancer and early cancer at endoscopy.

Further PLS-DA analysis demonstrates the diagnostic advantages of the integrated FP/HW Raman spectroscopy for early detection of gastric dysplasia. The complementary properties of the FP and HW Raman spectral modalities for enhancing tissue diagnosis can partially be explained by back-tracing the incorrect predictions of each Raman modality. The FP and HW Raman modalities made

incorrect predictions on 32 (normal: 23, dysplasia: 3, and adenocarcinoma: 6) and 46 (normal: 33, dysplasia: 3, and adenocarcinoma: 10) tissue sites, respectively, of the total 191 patients recruited. Integrated FP/HW Raman spectroscopy reduced the number of incorrect predictions to 22 (normal: 16, dysplasia: 2, and adenocarcinoma: 4). Further investigations show that the Raman signals of misclassified FP Raman spectra are extremely weak but with higher AF background and its resultant low FP spectral signal to noise ratios (SNR) $\sim 0.0118 \pm 0.0054$ (SNR: defined as the Raman intensity of normal gastric tissue Raman peak at 1445 cm^{-1} divided by its AF background); while the SNR (defined as the mean values of Raman intensity of normal gastric tissue Raman peak at 1445 cm^{-1} divided by its AF background, and the Raman intensity of normal gastric tissue Raman peak at 2940 cm^{-1} divided by its AF background) of the integrated FP/HW Raman spectroscopy is 0.0777 ± 0.0287 , which shows a significant enhancement through its addition of the HW Raman modality, and therefore, improving the performance of the tissue classification with integrated FP/HW Raman spectroscopy. In addition, the enhanced diagnostic performances of the FP/HW Raman spectroscopy were confirmed by its comparisons against the routinely used WLR endoscopy technique (Table 4). On per tissue site basis, the FP/HW Raman spectroscopy provides a diagnostic accuracy of 99.3% (sensitivity: 90.9% (10/11); specificity: 99.9% (394/396)), which are higher than that of the WLR endoscopy (accuracy: 52.1%; sensitivity: 90.9% (10/11); specificity: 51.0% (202/396)) (Table 4). Besides, a Netherland study⁵¹ reported a sensitivity of 51% and specificity of 67% for the diagnosis of premalignant lesions including gastric dysplasia using

WLR endoscopy, which are similar to the performances of WLR endoscopy in our study (Table 4), reconfirming the unrivaled advantages of the FP/HW Raman spectroscopy system developed for the diagnosis of gastric dysplasia. The real-time FP/HW fiber-optic Raman spectroscopy system developed is now routinely used by both experienced and novice endoscopists without difficulty at Endoscope Centre at National University Health System (NUHS), Singapore as an adjunct for screening and surveillance of gastric patients and targeted biopsies during clinical endoscopic examination. Some limitations of the current study should be pointed out. One notes that the current *in vivo* FP/HW tissue Raman datasets acquired are unbalanced with a relatively larger number of true negatives but a smaller pathologic tissue (e.g. dysplasia), which may give rise to a bias overall accuracy within the PLS-DA models. The patients recruited are not randomized and might have selection bias (i.e., non-responder bias). Large prospectively randomized multicenter clinical trials are underway for further assessing true clinical diagnostic value of the real-time FP/HW Raman spectroscopy in gastrointestinal examinations.

In summary, we demonstrate that high-quality *in vivo* gastric tissue Raman spectra can be acquired from gastric patients in real-time by using simultaneous FP/HW fiber-optic Raman spectroscopy technique. Significant FP/HW Raman spectral differences are observed among normal, dysplasia and adenocarcinoma of the stomach. The use of complementary biochemical/biomolecular information harvested through the integrated FP/HW Raman spectroscopy significantly enhances the diagnosis of gastric dysplasia as compared to either the FP or HW Raman technique

alone. Further comparison with the routinely used WLR endoscopy reconfirms the advantages of the integrated FP/HW Raman spectroscopy for gastric dysplasia diagnosis. This unique biomolecular endoscopic approach based on the simultaneous FP/HW fiber-optic Raman spectroscopy opens a new avenue for enhancing real-time, *in vivo* diagnosis of gastric precancer and early cancer at the molecular level during clinical endoscopy.

Acknowledgements

This project was supported by the National Medical Research Council (NMRC) (CIRG/1331/2012; BnB/0012c/2014; BnB13dec037), and the Academic Research Fund (AcRF)-Tier 2 from the Ministry of Education (MOE) (MOE2014-T2-1-010), Singapore.

References

1. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet - Tieulent and A. Jemal, *CA-Cancer J. Clin.*, 2015, 65, 87-108.
2. T. Gotoda, H. Yamamoto and R. M. Soetikno, *J. Gastroenterol.*, , 2006, 41, 929-942.
3. M. Rugge, F. Farinati, R. Baffa, F. Sonego, F. Di Mario, G. Leandro, F. Valiante and D. Interdisciplinary Group on Gastric Epithelial, *Gastroenterology*, 1994, 107, 1288-1296.
4. G. Y. Lauwers and R. H. Riddell, *Gut*, 1999, 45, 784-784.
5. M. Lansdown, P. Quirke, M. F. Dixon, A. T. Axon and D. Johnston, *Gut*, 1990, 31, 977-983.
6. B. Mayinger, M. Jordan, T. Horbach, P. Horner, C. Gerlach, S. Mueller, W. Hohenberger and E. G. Hahn, *Gastrointest Endosc.*, 2004, 59, 191-198.
7. G. Zonios, L. T. Perelman, V. Backman, R. Manoharan, M. Fitzmaurice, J. Van Dam and M. S. Feld, *App. Opt.*, 1999, 38, 6628-6637.
8. L. M. Almond, J. Hutchings, N. Shepherd, H. Barr, N. Stone and C. Kendall, *J. Biophotonics.*, 2011, 4, 685-695.
9. J. C. C. Day, R. Bennett, B. Smith, C. Kendall, J. Hutchings, G. M. Meaden, C. Born, S. Yu and N. Stone, *Phys. Med. Biol.* 2009, 54, 7077.
10. R. O. P. Draga, M. C. M. Grimbergen, P. L. M. Vijverberg, C. F. P. v. Swol, T. G. N. Jonges, J. A. Kummer and J. L. H. Ruud Bosch, *Anal.Chem.*, 2010, 82, 5993-5999.

11. Z. Huang, M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh and K. G. Yeoh, *J. Biomed. Opt.*, 2010, 15, 037017-037017-037015.
12. Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam and H. Zeng, *Int. J. Cancer*, 2003, 107, 1047-1052.
13. Z. Huang, S. K. Teh, W. Zheng, J. Mo, K. Lin, X. Shao, K. Y. Ho, M. Teh and K. G. Yeoh, *Opt. Lett.*, 2009, 34, 758-760.
14. Y. Komachi, H. Sato, K. Aizawa and H. Tashiro, *App. Opt.*, 2005, 44, 4722-4732.
15. J. T. Motz, M. Hunter, L. H. Galindo, J. A. Gardecki, J. R. Kramer, R. R. Dasari and M. S. Feld, *App. Opt.*, 2004, 43, 542-554.
16. M. G. Shim, B. C. Wilson, E. Marple and M. Wach, *Appl. Spectrosc.*, 1999, 53, 619-627.
17. M. A. Short, S. Lam, A. McWilliams, J. Zhao, H. Lui and H. Zeng, *Opt. Lett.*, 2008, 33, 711-713.
18. N. Stone, C. Kendall, J. Smith, P. Crow and H. Barr, *Faraday Discuss*, 2004, 126, 141-157.
19. S. K. Teh, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh and Z. Huang, *Brit. J. Cancer*, 2008, 98, 457-465.
20. C. A. Patil, I. J. Pence, C. A. Lieber and A. Mahadevan-Jansen, *Opt. Lett.*, 2014, 39, 303-306.
21. S. Koljenovic', T. C. Bakker Schut, R. Wolthuis, B. de Jong, L. Santos, P. J. Caspers, J. M. Kros and G. J. Puppels, *J. Biomed. Opt.*, 2005, 10,

- 031116-03111611.
22. K. Lin, D. L. P. Cheng and Z. Huang, *Biosens. Bioelectron.*, 2012, 35, 213-217.
 23. J. Mo, W. Zheng, J. J. H. Low, J. Ng, A. Ilancheran and Z. Huang, *Anal.Chem.*, 2009, 81, 8908-8915.
 24. L. F. Santos, R. Wolthuis, S. Koljenović, R. M. Almeida and G. J. Puppels, *Anal.Chem.*, 2005, 77, 6747-6752.
 25. M. S. Bergholt, K. Lin, J. Wang, W. Zheng, H. Xu, Q. Huang, J.-I. Ren, K. Y. Ho, M. Teh, S. Srivastava, B. Wong, K. G. Yeoh and Z. Huang, *J. Biophotonics.*, 2015, 9999, n/a-n/a.
 26. J. Wang, K. Lin, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh and Z. Huang, *Sci. Rep.*, 2015, 5.
 27. J. Wang, M. S. Bergholt, W. Zheng and Z. Huang, *Opt. Lett.*, 2013, 38, 2321-2323.
 28. S. Duraipandian, M. Sylvest Bergholt, W. Zheng, K. Yu Ho, M. Teh, K. Guan Yeoh, J. Bok Yan So, A. Shabbir and Z. Huang, *J. Biomed. Opt.*, 2012, 17, 081418-081418.
 29. C. V. K. Raman, K. S. , *Nature*, 1928, 121, 501-502.
 30. American National Standards Institute, *American National Standard for the Safe Use of Lasers. ANSI Standard 2136.1-1986*, Washington, D.C., 1986.
 31. J. H. Torres, M. Motamedi, J. A. Pearce and A. J. Welch, *App. Opt.*, 1993, 32, 597-606.

32. F. P. Incropera, *Fundamentals of heat and mass transfer*, John Wiley & Sons, 2011.
33. M. D. Keller, R. H. Wilson, M.-A. Mycek and A. Mahadevan-Jansen, *Appl. Spectrosc.*, 2010, 64, 607-614.
34. C. Holmer, K. S. Lehmann, J. Wanken, C. Reissfelder, A. Roggan, G. Mueller, H. J. Buhr and J.-P. Ritz, *J. Biomed. Opt.*, 2007, 12, 014025-014028.
35. S. Thomsen, *Photochem. Photobiol.*, 1991, 53, 825-835.
36. J. Carletta, *Lang. Comput.*, 1996, 22, 249-254.
37. J. A. Hanley and B. J. McNeil, *Radiology*, 1982, 143, 29-36.
38. E. R. DeLong, D. M. DeLong and D. L. Clarke-Pearson, *Biometrics*, 1988, 837-845.
39. R. Wolthuis, M. van Aken, K. Fountas, J. S. Robinson Jr, H. A. Bruining and G. J. Puppels, *Anal. Chem.*, 2001, 73, 3915-3920.
40. E. M. Barroso, R. W. H. Smits, T. C. Bakker Schut, I. ten Hove, J. A. Hardillo, E. B. Wolvius, R. J. Baatenburg de Jong, S. Koljenović and G. J. Puppels, *Anal. Chem.*, 2015, 87, 2419-2426.
41. K. Badizadegan, V. Backman, C. W. Boone, C. P. Crum, R. R. Dasari, I. Georgakoudi, K. Keefe, K. Munger, S. M. Shapshay and E. E. Sheets, *Faraday Discuss*, 2004, 126, 265-279.
42. B. S. Worthington, J. A. Syrotuck and S. I. Ahmed, *J. Nutr.*, 1978, 108, 1402-1411.
43. J. H. Hughes, C. J. Leigh, S. S. Raab, S. Y. Hook, M. B. Cohen and M. J.

- Suhrland, *Cancer Cytopatho.*, 1998, 84, 289-294.
44. I. Georgakoudi, B. C. Jacobson, M. G. Müller, E. E. Sheets, K. Badizadegan, D. L. Carr-Locke, C. P. Crum, C. W. Boone, R. R. Dasari, J. Van Dam and M. S. Feld, *Cancer Res.*, 2002, 62, 682-687.
45. P. Correa and M. B. Piazuelo, *J. Dig. Dis.*, 2012, 13, 2-9.
46. G. J. Thomas, B. Prescott and D. E. Olins, *Science*, 1977, 197, 385-388.
47. O. H. Olsen, M. R. Samuelson, S. B. Petersen and L. Nrskov, *Phys. Rev. A*, 1989, 39, 3130.
48. P. Correa, *Cancer Res.*, 1988, 48, 3554-3560.
49. T. Pohle, S. Brändlein, N. Ruoff, H. K. Müller-Hermelink and H. P. Vollmers, *Cancer Res.*, 2004, 64, 3900-3906.
50. S. Leikin, V. A. Parsegian, W. Yang and G. E. Walrafen, *P Natl Acad Sci USA.*, 1997, 94, 11312-11317.
51. L. G. Capelle, J. Haringsma, A. C. de Vries, E. W. Steyerberg, K. Biermann, H. van Dekken and E. J. Kuipers, *Digest. Dis. Sci.*, 2010, 55, 3442-3448.

Table 1. Confusion matrix and tissue site based diagnostics using FP/HW Raman spectroscopy together with PLS-DA and majority voting.

FP/HW Raman Prediction			
Histopathology	Normal	Dysplasia	Adenocarcinoma
Normal	380	0	16
Dysplasia	2	9	0
Adenocarcinoma	4	0	30
Sensitivity (%)	96.0	81.8	88.2
Specificity (%)	86.7	95.3	95.6
Accuracy (%)	95.0	99.5	95.5

Table 2. Confusion matrix and tissue site based diagnostics using FP Raman spectroscopy together with PLS-DA and majority voting.

FP/HW Raman Prediction			
Histopathology	Normal	Dysplasia	Adenocarcinoma
Normal	373	0	23
Dysplasia	3	8	0
Adenocarcinoma	6	0	28
Sensitivity (%)	94.2	72.7	82.4
Specificity (%)	80.0	93.3	93.6
Accuracy (%)	92.7	99.3	93.4

Table 3. Confusion matrix and tissue site based diagnostics using HW Raman spectroscopy together with PLS-DA and majority voting.

FP/HW Raman Prediction			
Histopathology	Normal	Dysplasia	Adenocarcinoma
Normal	363	0	33
Dysplasia	2	8	1
Adenocarcinoma	10	0	24
Sensitivity (%)	91.7	72.7	70.6
Specificity (%)	71.1	90.0	91.2
Accuracy (%)	89.8	99.3	90.0

Table 4. Comparisons of diagnostic performance for the differentiation between normal and dysplastic gastric tissues on tissue site basis using real-time fiber-optic FP/HW Raman spectroscopy and conventional WLR endoscopy.

Diagnostic performance comparisons			
	Sensitivity	Specificity	Accuracy
WLR Endoscopy	90.9% (10/11)	51.0% (202/396)	52.1% (212/407)
Raman Spectroscopy	90.9% (10/11)	99.5% (394/396)	99.3% (404/407)

Figure Captions

Fig 1. (a) The mean *in vivo* FP/HW Raman spectra ± 1 standard deviations (SD) of normal (n=5160), dysplasia (n=155) and adenocarcinoma (n=477) acquired from 441 tissue sites of 191 gastric patients during clinical endoscopy. (b) Difference spectra (i.e., normal - dysplasia, normal - adenocarcinoma, dysplasia - adenocarcinoma) ± 1 SD resolving the distinct spectral feature of gastric dysplasia.

Fig 2. (a) ANOVA of the three tissue categories over the entire spectral range (i.e., 800-1800 cm^{-1} , and 2800-3600 cm^{-1}). (b) Histogram ± 1 SD of the most diagnostic significant Raman peaks for the three tissues categories classifications (853 cm^{-1} , 1004 cm^{-1} , 1078 cm^{-1} , 1210 cm^{-1} , 1265 cm^{-1} , 1302 cm^{-1} , 1335 cm^{-1} , 1445 cm^{-1} , 1618 cm^{-1} , 1655 cm^{-1} , 2850 cm^{-1} , 2885 cm^{-1} , 2940 cm^{-1} , 3250 cm^{-1} , and 3300 cm^{-1}).

Fig 3. Representative hematoxylin and eosin (H&E)-stained histopathologic slides (original magnification, $\times 200$) corresponding to different gastric tissue types measured. (a) normal foveolar epithelium; (b) dysplastic epithelium, and (c) adenocarcinoma.

Fig 4. The first two diagnostically significant LVs of Raman spectral features for gastric tissue classification (i.e., normal, dysplasia and adenocarcinoma) algorithms development using partial least squares (PLS) – discriminant analysis (DA) and leave one patient out cross validation (LOPCV).

Fig 5. The posterior probabilities of 191 gastric patients belonging to normal stomach (n=396 sites), gastric dysplasia (n=11 sites) and adenocarcinoma (n=34 sites) based on the (a) FP, (b) HW and (c) the integrated FP/HW Raman techniques, respectively. (●) normal; (▲) dysplasia; and (■) adenocarcinoma.

Fig 6. Receiver operating characteristic (ROC) curves for separating gastric dysplasia from normal gastric tissue. The areas under the ROC curves (AUC) are 0.935, 0.850 and 0.982 for the FP, HW, and the integrated FP/HW Raman techniques, respectively.