REAL-TIME FIBER-OPTIC RAMAN SPECTROSCOPY FOR EARLY DIAGNOSIS OF PRECANCER AND CANCER IN VIVO IN THE UPPER GASTROINTESTINAL TRACT

MADS SYLVEST BERGHOLT

NATIONAL UNIVERSITY OF SINGAPORE

2014

REAL-TIME FIBER-OPTIC RAMAN SPECTROSCOPY FOR EARLY DIAGNOSIS OF PRECANCER AND CANCER IN VIVO IN THE UPPER GASTROINTESTINAL TRACT

REAL-TIME FIBER-OPTIC RAMAN SPECTROSCOPY FOR EARLY DIAGNOSIS OF PRECANCER AND CANCER IN VIVO IN THE UPPER GASTROINTESTINAL TRACT

MADS SYLVEST BERGHOLT

(MSc, Eng.)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOMEDICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have accordingly acknowledged all the sources of information, which have been used in the dissertation.

This dissertation has not been submitted for any degree in any university previously.

_____

MADS SYLVEST BERGHOLT

17 January 2014

Dedicated to my family and friends

Acknowledgements

It is a great pleasure to thank those who were involved in the completion of this PhD dissertation. First and foremost, I would like to express my solemn gratitude to scientific supervisor Prof. Zhiwei Huang for providing me the opportunity to pursue the PhD in his research group. I like to thank Prof. Huang for his insightful technical discussions, continuous encouragement and feedback throughout this PhD project. It has been a privilege to have Prof. Zhiwei Huang as a supervisor throughout the project and I will express my special thanks for his funding support and prompt actions that enabled the multiunit clinical translational project to rapidly progress. I would also like to acknowledge our clinical collaborators who are involved in this project, in particular Prof. Khek Yu Ho, Assoc. Prof. Ming Teh, Assoc. Prof. Khay Guan Yeoh, Assoc. Prof. Jimmy So and all the doctors, research nurses and nurses from Endoscopy Center and Gastrotranslational Research Office, National University Hospital, Singapore. I would further like to thank my colleagues at Optical Bioimaging Laboratory, Department of Biomedical Engineering: Dr. Wei Zheng, Dr. Shiyamala Duraipandian, Dr. Lin Kan, Mr. Wang Jianfeng, Dr. Lin Jian, Mr. Wang Zi, Mr. Arnold Benjamin, Mr. Teh Seng Khoon, Dr. Mo Jianhua, Dr. Shao Xiaozhuo, Dr. Lu Fake, for their support and collaborations. I would also like to express my solemn appreciation to my beloved parents, brothers and friends. Finally, I would like to acknowledge the study funding agencies: the National Medical Research Council, the Biomedical Research Council, and the Academic Research Fund from the Ministry of Education, Singapore.

Publications (Peer-Reviewed Journals)

1. M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, A. Shabbir, Z. Huang, "Fiber-optic confocal Raman spectroscopy for real-time *in vivo* diagnosis of dysplasia in Barrett's esophagus," Gastroenterology, 146(1), 27-32 (2014)**.** Impact factor 12.835.
2. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, A. Shabbir, Z. Huang ” Fiber-optic Raman spectroscopy probes gastric carcinogenesis *in vivo* at endoscopy” J. Biophotonics 6(1), 49–59 (2013). Impact factor 3.099.
3. M. S. Bergholt, W. Zheng, Z. Huang, “Development of a multiplexing fingerprint and high wavenumber Raman spectroscopy technique for real-time *in vivo* tissue Raman measurements at endoscopy,” J. Biomed. Opt. 18(3), 30502 (2013). Impact factor 3.145.
4. S. Duraipandian, M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Yan So, and Z. Huang, "Real-time Raman spectroscopy for *in vivo*, online gastric cancer diagnosis during clinical endoscopic examination," J. Biomed. Opt. 17(8), 081418 (2012). Impact factor 3.145.
5. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang ”Characterizing variability in *in vivo* Raman spectra of different anatomical sites in the upper gastrointestinal tract toward cancer detection,” J. Biomed. Opt. 16(3), 037003 (2011). Impact factor 3.145.

Publications (Conference proceedings)

1. M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So Z. Huang ”Real-time depth-resolved fiber-optic Raman endoscopy for *in vivo* diagnosis of gastric precancer” Proceedings of SPIE ‘accepted’ (2014).

2. S. Duraipandian, M. S. Bergholt, W. Zheng, Z. Huang ”Quantitative fiber optic Raman spectroscopy for tissue Raman measurements” Proceedings of SPIE ‘accepted’ (2014).

3. Z. Huang, M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, A. Shabbir, ” Moving Raman spectroscopy into real-time, online diagnosis and detection of precancer and cancer *in vivo* in the upper GI during clinical endoscopic examination” Proceedings of SPIE, 8577, 85770G (2013).

4. M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So Z. Huang ” Real-time depth-resolved Raman endoscopy for in vivo diagnosis of dysplasia in Barrett's esophagus” Proceedings of SPIE, 8576, 85760D (2013).

5. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang ”Detection of malignant lesions *in vivo* in the upper gastrointestinal tract using image-guided Raman endoscopy” Proceedings of SPIE, 8219, 82190M (2012).

6. M. S. Bergholt, W. Zheng, K. Lin, K. Yu Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang ”Multimodal endoscopic imaging and Raman spectroscopy for improving *in vivo* diagnosis of gastric malignancies during clinical gastroscopy” Proceedings of SPIE, 7990, 79900A (2011).

7. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang ”Image-guided real-time Raman endoscopy for *in vivo* diagnosis of gastric precancer and cancer during clinical gastroscopy” DDW, Chicago Illinois, (2011).

8. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, Z. Huang ”*In vivo* Raman spectroscopy integrated with multimodal endoscopic imaging techniques for improving the early diagnosis of gastric malignancies” Proceedings of SPIE, 7560, 756003 (2010).

Publications (Conference abstracts)

1. S. Duraipandian, M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, A. Shabbir, Z. "Image-guided fiber optic confocal Raman spectroscopy for real-time diagnosis of intestinal metaplasia *in vivo* at endoscopy", SGCC 2013, National University Hospital of Singapore, Singapore.

2. J. Wang, M. S. Bergholt, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, A. Shabbir, Z. "Development of a confocal Raman endoscopic probe for sensitive detection of gastric dysplasia *in vivo*", SGCC 2013, National University Hospital of Singapore, Singapore.

3. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, A. Shabbir, Z. Huang "Image-guided Raman spectroscopy for real-time *in vivo* diagnosis of Barrett's esophagus during endoscopic examination", DDW 2012, San Diego, United States.

4. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang, "Raman endoscopy for i*n vivo* diagnosis of Barrett's esophagus", SHBC 2011, SUNTEC Singapore International Convention & Exhibition Centre, Singapore (Best Poster Award).

5. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang "On-line image-guided Raman endoscopy for *in vivo* diagnosis of gastric cancer", APDW 2011, SUNTEC Singapore International Convention & Exhibition Centre, Singapore.

6. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang "On-line *in vivo* diagnosis of gastric and esophageal neoplasia using

image-guided Raman endoscopy", SGCC 2011, Shaw Foundation Alumni House National University of Singapore, Singapore.

7. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang "Image-guided real-time Raman endoscopy for *in vivo* diagnosis of gastric precancer and cancer during clinical gastroscopy", DDW 2011, Chicago Illinois, United States.
8. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, Z. Huang" Raman endoscopy for *in vivo* diagnosis of gastric precancer and cancer at gastroscopy", SHBC 2010, SUNTEC Singapore International Convention & Exhibition Centre, Singapore.
9. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang "Image-guided Raman endoscopy for *in vivo* differentiation between malignant and benign ulcers in the gastric", APDW 2010, Kuala Lumpur Convention Centre, Malaysia.
10. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, Z. Huang "Raman endoscopy for *in vivo* diagnosis of gastric cancer", SGCC 2010, Shaw Foundation Alumni House, National University of Singapore, Singapore.
11. Z. Huang, M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, "Image-guided Raman spectroscopy for *in vivo* diagnosis of gastric precancer at gastroscopy", ICORS 2010, Boston, United States.

Table of Contents

Abstract

Gastrointestinal (GI) malignancies including esophagus and gastric cancer are diseases with high mortality rates. The patients suffering from GI cancers have poor survival rates mainly due to the advanced stages upon the initial time of diagnosis. Early diagnosis and localization of premalignant and malignant lesions together with appropriate curative treatment are therefore critical measures to reducing the mortality rates of the patients. The identification of precancer and early flat cancerous lesions can however be very challenging for clinicians using conventional endoscopy such as white light reflectance (WLR), narrowband imaging (NBI) and autofluorescence imaging (AFI). Current diagnostic gold standard relies on endoscopic biopsies that are invasive and a major burden to both the patient and society.

Raman spectroscopy is an inelastic light scattering technique with high molecular specificity. For tissue and cells, Raman spectroscopy is able to harvest a wealth of information from the myriad of inter- and/or intra- cellular components (i.e., proteins, lipids and deoxyribonucleic acid (DNA)). The main objective of this dissertation is to develop a fiber-optic Raman spectroscopy technique for real-time *in vivo* diagnosis of precancer and cancer in the upper GI during endoscopy.

We develop a novel trimodal image-guided (i.e., WLR, NBI and AFI) fiber-optic rapid Raman spectroscopy technique for real-time *in vivo* tissue Raman measurements at clinical endoscopic examinations. The novel Raman endoscopy technique enables high-quality *in vivo* Raman spectra to be measured within 0.5 s under the guidance of wide-field endoscopic imaging modalities (i.e., WLR/NBI/AFI). We utilize the novel Raman endoscopy technique for characterization of tissues belonging to different anatomical regions in the upper GI tract *in vivo* and explore the implications for early detection of neoplastic lesions. We

show that the inter-anatomical variability of Raman spectra among normal gastric tissue (antrum, body, and cardia) is subtle compared to cancerous tissue transformation. *In vivo* Raman spectroscopy reveals the progressive changes of biochemical constituents in gastric tissue associated with preneoplastic and neoplastic transformation (i.e., intestinal metaplasia (IM), dysplasia and adenocarcinoma). Multi-class probabilistic partial least squares-discriminant analysis (PLS-DA) diagnostic algorithms based on *in vivo* Raman spectra are able to identify IM with sensitivity of 46.67% and specificity of 87.55%; dysplasia with sensitivity of 83.33%; specificity of 95.80%, and adenocarcinoma with sensitivity of 84.91% and specificity 95.57%, respectively based on 1277 *in vivo* Raman spectra from 83 gastric patients.

In order to realize prospective clinical diagnostics we develop an on-line diagnostic framework and demonstrate for the first time the prospective cancer diagnosis in 10 gastric patients at gastroscopy. Further, to significantly improve the sensitivity for epithelial precancerous lesions, a novel fiber-optic confocal Raman endoscopy technique is presented using a bevelled fiber-optic Raman probe coupled with a sapphire ball lens. We apply the online fiber-optic confocal Raman technique prospectively in 77 patients. *In vivo* confocal Raman spectra show significantly spectral differences ($p < 0.001$) among different esophageal tissue types (i.e., columnar-lined epithelium, non-dysplastic BE and dysplastic BE). Fiber-optic confocal Raman spectroscopy provides a diagnostic sensitivity of 87.0% (67/77) and a specificity of 84.7% (610/720) for *in vivo* detection of dysplasia in 77 BE patients prospectively based on the endogenous biomolecular signatures.

Finally we develop a second generation multiplexed Raman endoscopy technique using a single laser light together with a volume phase holographic (VPH) grating that simultaneously acquires both fingerprint (FP) and high wavenumber

(HW) tissue Raman spectra in high resolution. We utilize a customized VPH dual-transmission grating, which disperses the incident Raman scattered light vertically onto two separate segments (i.e., 150 to 1950 cm^{-1}; 1750 to 3600 cm^{-1}) of a charge-coupled device camera (CCD). The rapid multiplexed Raman spectroscopy technique covering both FP and HW ranges developed has great potential for further improving *in vivo* tissue diagnosis and characterization at endoscopy.

List of Tables

List of Figures

List of Abbreviations

AFI	Autofluorescence imaging
AGA	American Gastrointestinal Association
a/LCI	Angle-resolved low-coherence interferometry
ANOVA	Analysis of variance
APU	Auxiliary power unit
AUC	Area under curve
BE	Barrett's esophagus
CART	Classification and regression threes
CCD	Charge coupled device
CI	Confidence interval
CIS	Carcinoma in situ
CLE	Confocal laser endomicroscopy
CT	Computed tomography
DA	Discriminant analysis
DNA	Deoxyribonucleic acid
DOT	Diffuse optical tomography
EAC	Esophageal adenocarcinoma
eCLE	Endoscopic based confocal laser endomicroscopy
EMR	Endoscopic mucosal resection
EMSC	Extended multiplicative scatter correction
ESD	Endoscopic submucosal dissection
ESS	Elastic scattering spectroscopy
FBG	Fiber Bragg grating
FOV	Field of view
FP	Fingerprint
GCEP	Gastric Cancer Epidemiology, Clinical and Genetics Programme
GCP	Good clinical practice
GE	Gastroesophageal
GERD	Gastroesophageal reflux disease
GRIN	Gradient index lens
GUI	Graphical user interface
H&E	Hematoxylin and eosin
HP	Helicobacter Pylori
HW	High-wavenumber
ICH	International Conference on Harmonization
IM	Intestinal metaplasia
IRB	Institutional review board
KHD	Kramer-Heisenberg dispersion relation
LGD	Low-grade dysplasia
LSD	Least significant difference

LV	Latent variable
MCSMF	Multi-core single-mode fibers
MRI	Magnetic resonance imaging
MSC	Multiple scatter correction
NADH	Nicotinamide adenine dinucleotide
NBI	Narrowband imaging
NHG	National Healthcare Group
NIR	Near-infrared
NNLSM	Non-negativity least squares minimization
NPV	Negative predictive value
NUHS	National University Hospital, Singapore
NUS	National University of Singapore
OCT	Optical coherence tomography
pCLE	Probe based confocal laser endomicroscopy
PCA	Principal component analysis
PC	Principal component/personal computer
PLS-DA	Partial least squares – discriminant analysis.
PLS-GA	Partial least squares – genetic algorithms.
PMT	Photo multiplier tube
PPV	Positive predictive value
QE	Quantum efficiency
RNA	Ribonucleic acid
ROC	Receiver operating characteristics
SD	Standard deviation
SE	Standard error
SNR	Signal to noise ratio
SR	Supplementary Raman
SSRS	Shifted subtracted Raman spectroscopy
SVM	Support vector machines
TURBT	Transurethral resection of bladder tumors
VPH	Volume phase holographic
WLR	White light reflectance
Nd:YAG	Neodymium-doped yttrium aluminum garnet

Chapter 1. Introduction

1.1 Background

1.1.1 Gastric cancer

Gastric cancers remain a serious disease and have one of the worst 5-year survival rates among malignancies. A total of 989,600 new stomach cancer cases and 738,000 deaths are estimated to occur annually world-wide, accounting for 8% of the total cancer cases and 10% of total deaths.[1, 2] In Singapore, gastric cancer is also a serious health burden. Gastric cancer is the fifth most common cancer in males and seventh most common cancer in females in Singapore.[3] The age-standardized rate of gastric cancer is 21.4 per 100 000 per year in Chinese males and 10.8 per 100,000 per year in Chinese females.[3] The epidemiology of gastric cancer is characteristic with significant demographic differences.[4] The incidence is particularly high in East Asia, Eastern Europe, Russia and parts of Central and South America, and it is approximately twice as high among men than among women.[4] In East-Asia (including China, Taiwan, Japan, Korea) and Russia the incidence of gastric cancer is much higher than other parts of the world. In multiracial countries such as Malaysia and Singapore, gastric cancer is more common in Chinese people than in those of Malay and Indian origin. The cause of gastric cancer is not fully understood but is likely linked with both genetic and environmental factors such as smoking, diet (smoked foods, salted fish and meat, and pickled vegetables) and Helicobacter Pylori (HP) infection.[5] Infection by the bacterium HP is believed to be the primary cause of most intestinal-type stomach cancers while autoimmune atrophic gastritis, intestinal metaplasia (IM) and various genetic factors are associated with increased risk levels and is considered the part of the intestinal type carcinogenesis pathway.[6] Epidemiological data have shown that subjects who test positive for HP have at least

two-times increased risk of developing gastric cancer compared with those who test negative.[7] HP eradication is therefore considered an important parameter to reduce the incidence of gastric cancer and extensive efforts for HP eradication have therefore been initiated.[8]

Gastric adenocarcinomas are the most common cancer type in the gastric and can be divided into two different subtypes (Lauren classification), that is: intestinal type gastric cancer and diffuse type gastric cancer. A higher incidence of intestinal type gastric cancer has been observed in males, blacks and older subjects, while the diffuse type gastric cancer has a similar incidence in both genders and is more frequent in younger individuals.[6, 9] Intestinal type gastric carcinogenesis is widely accepted to be a multistep cascade of pathological events and has been described with the Correa model.[6] The Correa pathway is depicted in Figure 1.1, which shows that the disease transformation occurs in sequential steps after exposure to HP, that is: chronic gastritis, atrophy, IM, dysplasia and ultimately neoplasia. The intestinal-type tumor is histopathologically characterized by irregular tubular structures, harboring pluristratification, multiple lumens, reduced stroma. Often, these tumors associate with severe IM in neighboring mucosa.

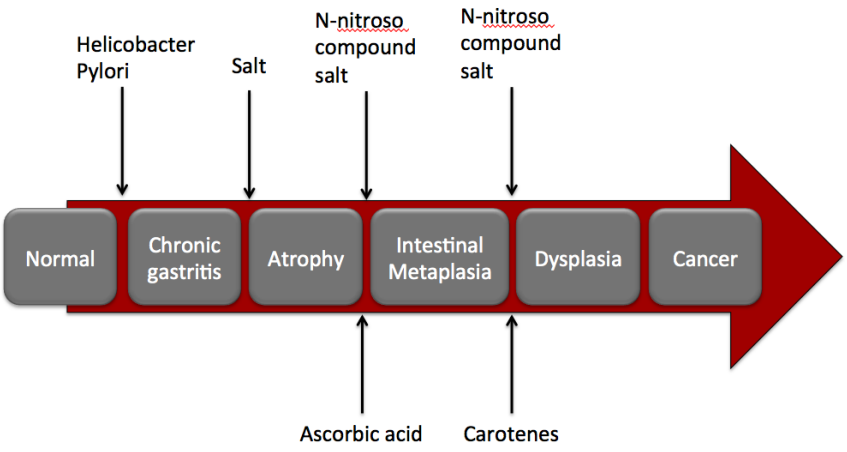


Figure 1.1 Schematics of the Correa model of sequential mucosal changes in the development of intestinal type gastric cancer on the basis of Helicobacter pylori (HP) driven chronic active gastritis.

Diffuse type gastric cancer on the other hand is considered more dependent on the genetic profile of the individuals.[10] Unlike the complex disease pathway that characterizes intestinal-type carcinogenesis (Figure 1.1), diffuse gastric adenocarcinomas display prominent molecular abnormalities including defective intercellular adhesions. In most cases, this results from loss of expression of the cell adhesion protein E-cadherin.[11] The tumor is histopathologically characterized by tumor cells that are discohesive and secrete mucus, which is delivered in the interstitium, producing large pools of mucus/colloid. The diffuse type cancers, are highly metastatic and characterized by more rapid disease progression than intestinal type gastric cancers.[12] Partial or full gastric resections are commonly performed for these cancer types to reduce the risk of recurrence in the patients.

1.1.2 Esophageal cancer

Esophageal cancers can be divided into squamous cell carcinoma and esophageal adenocarcinoma. Both diseases have high mortality rates since the disease is often discovered late when the patient is symptomatic (e.g., dysphagia). This dissertation will mainly focus on Barrett's esophagus (BE) associated adenocarcinoma in the distal esophagus since this remains a major endoscopic challenge. BE is a complication of gastroesophageal (GE) reflux disease (GERD) and is a precursor of esophageal adenocarcinoma (EAC), which occurs above the proper gastroesophageal junction.[13] Since the late 1970s, the BE disease has had a significant rise in incidence rates (>350%) in the developed countries.[13, 14] Prevalences for BE ranging from 5 to 25% have been reported, and approximately 5-15% of patients undergoing endoscopy for reflux symptoms have been found to be associated with BE.[15, 16] BE is a

metaplastic precursor of EAC. There is some confusion in the literature regarding the exact definition of BE. In this dissertation we will follow American Gastrointestinal Association (AGA) recommendations with BE defined as the presence of goblet cells (i.e., IM phenotype) in the proper esophagus. Figure 1.2 shows the pathological transformation in BE. BE carcinogenesis is thought to develop from esophagitis, IM, dysplasia and ultimately cancer.[17] The precise mechanisms through which reflux constituents drive the carcinogenic process in the esophagus are however poorly understood. GERD might act individually or together at many points in the disease pathway, from the development of BE through to progression from dysplasia to EAC.[18] Inflammatory processes associated with GERD are believed have a key role in disease progression.[18]

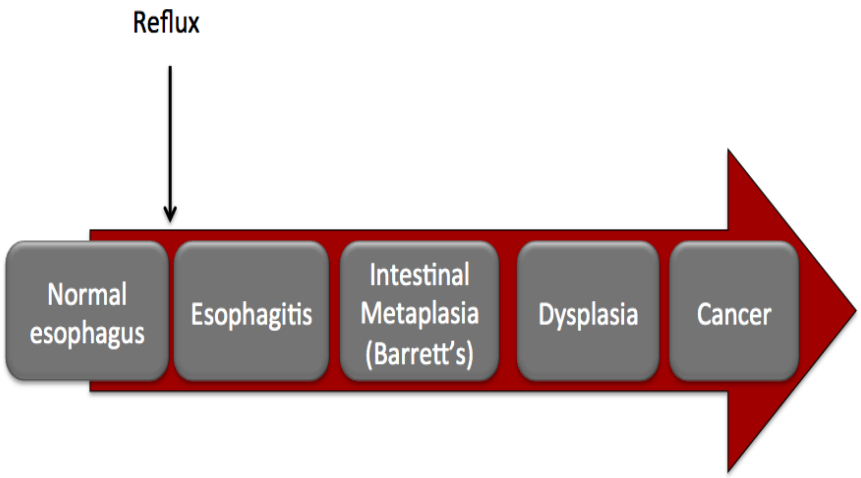


Figure 1.2 The Barrett's esophagus (BE) carcinogenesis sequence associated with gastroesophageal reflux disease (GERD).

1.1.3 Conventional cancer screening methods

Flexible endoscopy in patients with suspicious symptoms in the upper GI is the primary choice for the clinician. Current standard flexible endoscopic diagnostics in the esophagus and gastric are based on conventional white-light reflectance (WLR) endoscopic imaging. In recent years, optical imaging methods, such as the narrow-

band imaging (NBI) technique, which enhances visualization of irregular mucosal and vascular patterns, has shown promising diagnostic potential for *in vivo* detection of precancer and early cancer at endoscopy.[19] While NBI imaging techniques provide higher detection sensitivities, this wide-field endoscopic imaging modality still suffer from moderate diagnostic specificities due to inter-observer dependence and the inherent lack of ability to reveal endogenous biochemical information about the tissue.[20] Very recently, autofluorescence imaging (AFI) has emerged as a diagnostic tool.[21] AFI reveals biochemical information by imaging the endogenous fluorephores in the tissue but this technique has not yet gained widespread acceptance. Endoscopic biopsies of suspicious lesions that are sent for histopathological examination remain the current gold standard in the upper GI.

1.1.4 Motivation and research objectives

Early identification and intervention are the most critical measures to reducing the cancer-related mortality rates in the GI tract.[8] Dysplastic lesions or grossly inconspicuous cancers are however endoscopically indistinguishable from the surrounding benign tissue using conventional WLR/NBI/AFI endoscopy. Subtle changes of early lesions (e.g., dysplasia, carcinoma in situ (CIS)) may not be apparent, limiting the diagnostic sensitivity. This is because conventional endoscopic modalities heavily rely on visual assessment of structural and morphological details and provide little or no biomolecular information. Endoscopic inspections are therefore associated with a vast number of unnecessary biopsies, which are clinically labour intensive and also a burden to the patients and society. On top of this, histopathological characterization of minute specimens can also be highly subjective. For instance, inter-observer disagreements of up to >50% have been reported for low-

grade dysplasia (LGD) in BE.[22, 23] The identification of surgical margins remains another great challenge in clinical endoscopy. Achieving negative resection margins, defined as the absence of dysplasia or cancer, provides lower associated rates of local recurrence. Residual lesions after resection however frequently occur.[24] In clinical practice, as many as ~20-30% of patients undergoing endoscopic mucosal resection (EMR) require a second surgical procedure due to local recurrence indicative of incomplete cancer resection that were missed on margin assessment.[24] Thus, it is of imperative clinical value to develop non-invasive and sensitive optical diagnostic technologies that can assist in guiding endoscopists for the targeted biopsies of suspicious lesions (i.e., precancer and cancer), and for surgery and management of GI patients.[3] The development of a rapid and non-invasive or minimally invasive optical point-wise diagnostic technique enabling the direct assessment of specific biochemical and biomolecular information about suspicious lesions *in vivo* to complement the WLR/AFI/NBI imaging would represent a significant advance in the early diagnosis and management of upper GI malignancies. Reflectance spectroscopy, fluorescence and Raman spectroscopy have become promising candidates for *in vivo* diagnostics in medicine. Among these optical spectroscopic techniques, Raman spectroscopy is the most biomolecular-specific but also the one that is most technically difficult to realize in practice. Raman spectroscopy is an inelastic light scattering technique that is capable of probing the highly specific vibrational frequencies of biomolecules.[25] For tissues and cells, Raman spectroscopy enlightens the surface and subsurface cellular structures and conformations. Accumulating evidence has confirmed the capability of Raman spectroscopy to detect precancer and cancerous lesions *ex vivo* for different tissue types and organs.[26-31] For instance, the diagnostic sensitivities and specificities in the range of ~85–95% and ~90–98% have

been reported for differentiation between different pathologic types (e.g., IM, dysplasia and adenocarcinoma) of gastric tissue using NIR Raman spectroscopy associated with multivariate analysis.[26-30] Recently the first commercial Raman device has been developed (Aura, Verisante Inc). To date there however does not exist Raman techniques that routinely can be applied in the gastrointestinal tract.

1.2 Specific aims and organization

The main objective of this dissertation is to develop a real-time fiber-optic Raman spectroscopy technique for early diagnosis of precancer and cancer in the upper GI tract *in vivo*. The specific aims are formulated as follows:

1. To develop a real-time fiber-optic Raman spectroscopy technique for the upper GI tract and characterize/explore inter-anatomical Raman spectral variability of normal tissue in the esophagus and gastric and compare with that of cancerous transformation *in vivo* (Chapter 4).
2. To explore the diagnostic utility of fiber-optic Raman spectroscopy for elucidation of gastric intestinal type carcinogenesis and diagnosis of different precancerous and cancerous lesions *in vivo* at endoscopy (Chapter 5).
3. To develop an on-line fiber-optic Raman spectroscopic framework in order to realize real-time prospective *in vivo* tissue diagnosis during clinical endoscopy (Chapter 6).
4. To improve the sensitivity to epithelial lesions by developing a novel fiber-optic confocal Raman spectroscopy technique and demonstrate the prospective on-line *in vivo* tissue diagnosis of precancer in Barrett's esophagus (Chapter 7).

5. To develop a second-generation real-time Raman endoscopy technique that acquires both the fingerprint (FP) and high-wavenumber (HW) Raman signals simultaneously in high resolution for improving tissue characterization and diagnosis (Chapter 8).

Chapter 9 gives a discussion on the study limitations and scope of potential clinical applications of the fiber-optic Raman endoscope technique. Finally the thesis is concluded and future directions are given for further exploration of fiber-optic Raman spectroscopy in GI precancer and cancer detection and diagnosis.

Chapter 2. Overview of endoscopic imaging and spectroscopy techniques

2.1 Conventional endoscopic imaging

Flexible WLR endoscopy remains the current standard tool for diagnostics and is based on the reflection of visible broadband light from a xenon lamp (Figure 2.1). The diagnostic accuracy for identification of precancer and early flat cancer is however very limited. In one study based on 43 patients the sensitivity, specificity, positive and negative predictive values for detection of gastric premalignant lesions were 51, 67, 62 and 55% for WLR.[20] This is due to the fact that WLR only reveals gross morphological changes with no specific biochemical information about the tissue. WLR endoscopy is in general subjective and heavily depends on the experience of the endoscopist. Recently, NBI modality has emerged as a novel technique that enhances microvascular patterns (Figure 2.1).[19] NBI uses two discrete bands of light: One blue e.g., at 415 nm and one green at e.g., 540 nm. Since tissue have different absorption properties at the two wavelengths, narrowband blue light displays superficial capillary networks, while green light displays sub-epithelial vessels and when combined offer a high-contrast image of the tissue surface. Specificity, sensitivity, positive and negative predictive values for detection of premalignant lesions of 71, 58, 65 and 65% for NBI and 51, 67, 62 and 55% for WLE, have been reported respectively.[20] NBI improves the specificity compared to WLR but still remains in low end ~70%.[20]

Despite developments of zoom endoscopy and high-resolution/high-definition imaging, both WLR and NBI remain suboptimal for accurate *in vivo* detection of

precancers due to lack of revealing the cellular biochemical/biomolecular information. For this reason, AFI modality has been developed (Figure 2.1). AFI reveals more biochemical information through imaging of the endogenous fluorephores in the tissue. In endoscopic AFI mode, a rotating color filter wheel in front of the xenon light source sequentially generates blue light (390-470 nm) and green light (540-560 nm) for tissue illumination. A filter situated in front of the AFI CCD blocks the blue light excitation but enables tissue autofluorescence (500-630 nm) and reflected green light to filter through. The sequentially captured images of autofluorescence and green reflectance are integrated by the video processor into a real-time pseudocolor image in which normal mucosa appears green, and dysplastic and neoplastic tissue appears dark purple (Figure 2.1).

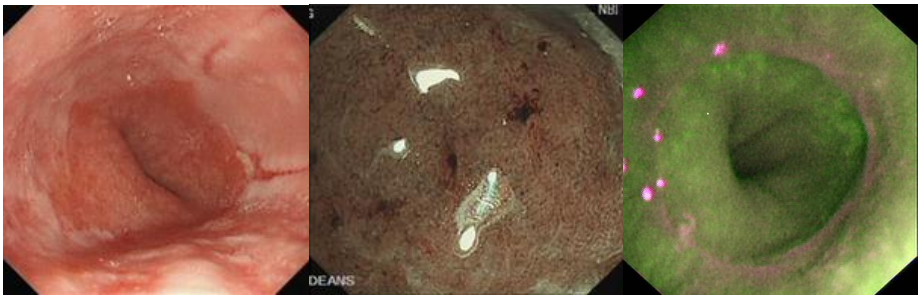


Figure 2.1 White light reflectance (WLR), narrowband imaging (NBI) and autofluorescence imaging (AFI). The images were obtained from Endoscopy Center with permission, National University Hospital of Singapore (NUHS).

The diagnostic utility of AFI technology however still remains debatable. One study compared WLR with AFI for experienced and inexperienced endoscopists.[21] The sensitivity, specificity, and accuracy were reported as 70%, 78%, and 75%, respectively, for AFI and 81%, 76%, and 78%, respectively, for WLE in the experienced group. In the less experienced group, sensitivity, specificity and accuracy were 80%, 81% and 80%, respectively, for AFI and 65%, 77%, and 72%, respectively, for WLE.[21] Very recently the clinical value of trimodal endoscopic imaging (i.e.,

WLR/NBI/AFI) has been investigated.[32] Results have indicated that AFI/NBI identified significantly more patients with IM than WLR alone.[32] In general, much research has been conducted in endoscopic imaging modalities, but recent directions have been guided more toward molecular imaging. Techniques such as near-infrared (NIR) labeled fluorescence markers (i.e., antibodies or peptides) and nanoparticles with integrated drug delivery capability are currently being investigated.[33] Since these technologies rely on exogenous agents and compounds, the translation into clinical medicine is more challenging.

2.2 Advanced endoscopic techniques

Confocal laser endomicroscopy

Confocal laser endomicroscopy (CLE) is a recently developed microscopic imaging modality with small field of view (FOV) for endoscopic applications.[34] There are two types of CLE: endoscope-based CLE (eCLE), which is integrated in the distal tip of the endoscope, and probe-based CLE (pCLE), a fiber-optic technique using the accessory channel of the endoscope. This technique uses intravenous agents (i.e., fluorescein) for improving the tissue contrast. CLE magnifies the tissue surface using a distal optical system for visualization of cellular structures (i.e., crypts, microvasculature etc.). Clinical data of CLE have been reported in GI diseases including BE, gastric, and colon.[34, 35] Promising results have been reported for detection of precancer and cancer.[71, 97] Figure 2.2 displays the *in vivo* confocal endomicroscopic image acquired at Endoscopy Center, National University Hospital of Singapore (NUHS), showing apparent presence of intestinal-type goblet cells (black spots) in BE. Although promising results have been reported, the accuracy and

clinical merit of this modality still remains debatable as indicated in a large randomized multicenter trial.[35] This technique needs expert training and is still associated with subjective interpretations.

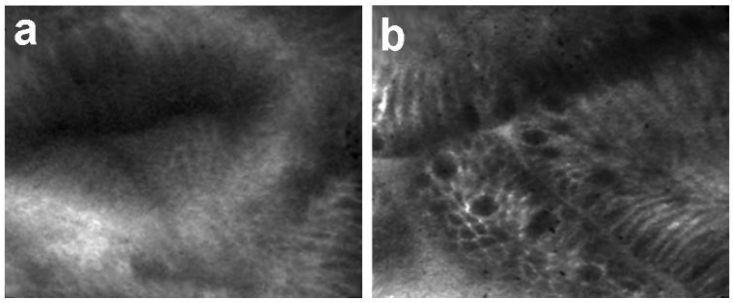


Figure 2.2 Confocal laser endomicroscopic images acquired from (a) normal and (b) Barrett's esophagus at Endoscopy Center, National University Hospital, Singapore (NUHS). (Pentax inc. Japan).

Optical coherence tomography

Optical coherence tomography (OCT) is another technique for morphological and architectural cross sectional imaging based on interferometry.[36] OCT obtains cellular images of epithelial and sub-epithelial tissue with a definition of 1 – 10 mm[37]. High levels of diagnostic accuracies have been reported (~80%) for differentiating between dysplasia and adenocarcinoma.[38] Despite encouraging results, the technique has not found its way into routine clinical endoscopy. Like CLE, this technique relies on subjective interpretations of gross architectural features of mucosa. This cross sectional technique could have potential use staging of precancer and cancers (i.e., assessment of tumor invasion). More exotic OCT techniques include angle-resolved low coherence interferometry (a/LCI) that is under development.[39]

Point-wise endogenous spectroscopic techniques

Optical spectroscopic techniques such as elastic light scattering spectroscopy (ESS), fluorescence and Raman spectroscopy have been comprehensively investigated for the evaluation of malignancies in tissues.[40-46] ESS reveals changes in

the physical properties of cells (i.e., nuclear size, shape and density and general morphology).[47] In 2006 Lovat and colleagues published a study showing high sensitivity of 92% however with low specificity of 60% for diagnosis in BE.[47] The technique has recently been extended to a rotational scanning device for the distal esophagus for Barrett's disease.[48] The clinical merit and diagnostic yield of this technique have not been reported to date and the technique has therefore not gained widespread acceptance yet.

Fluorescence spectroscopy is another point-wise technique but with higher biochemical specificity than ESS.[49] Fluorescence spectroscopy probes tissue composition by exciting biomolecules to an electronic eigenstate followed by relaxation to the electronic ground state. Fluorescence spectroscopy typically uses violet-blue excitation.[50] It's widely accepted that this technique probes fluorophores such as NADH, collagen, elastin, tryptophan etc. that undergo quantitative changes during tissue carcinogenesis.[50] Recently the use of NIR-excited autofluorescence has been proposed for diagnostic purposes but the *in vivo* utility has not been reported to date.[51-53]

Raman spectroscopy is the most molecular specific technique for endogenous tissue compositional analysis[26-31] but also the technique that is most challenging to realize in clinic settings.[54] *In vivo* Raman clinical endoscopic applications are extremely challenging and have been limited not only by the difficulty in capturing inherently weak tissue Raman signals, but also by the slow speed of spectral measurements (>5 sec). The miniaturization of flexible fiber-optic Raman probes that can pass down the accessory channel of medical endoscopes for effective tissue Raman light collections presents another great challenge in clinical Raman spectroscopy. To overcome the weak signals, more exotic techniques such as surface

enhanced Raman spectroscopy (SERS) are being explored *ex vivo*.[55] Since conventional NIR Raman spectroscopy is the main topic of this dissertation, the following sections introduces the more comprehensive physical and instrumental foundations of the Raman scattering effect as well as literature studies.

2.3 Raman spectroscopy

2.3.1 The Raman scattering effect

Raman spectroscopy is a non-destructive inelastic light scattering technique capable of probing vibrational modes of molecules without labeling. In 1928, the Indian physicist Sir Chandrasekhara Venkata Raman (1888-1970) discovered the Raman effect, and won the 1930 Nobel Prize in Physics "as the first Asian for his work on the scattering of light and for the discovery of the effect named after him".[56] When photons interact with a molecule, energy can be exchanged either through absorption or elastic/inelastic scattering. For an absorption process to happen, a change in net dipole must occur. In contrast, a net polarization change of the molecule is required for inelastic scattering to occur. The induced dipole moment P is proportional to the electric field of the incident wave E and the polarizability.

$$P = \alpha E \qquad [1]$$

Expanding this and using harmonic vibrations i.e., 3N-6 normal vibrations (or 3N-5 for linear molecules), one obtains

$$P = \alpha_0 E_0 \cos 2\pi \nu_0 t + \frac{1}{2}(\partial\alpha / \partial q)_0 q_0 E_0 [\cos\{2\pi(\nu_m + \nu_0)t\} + \cos\{2\pi(\nu_m - \nu_0)t\}] \qquad [2]$$

Where α_0 is the polarizability of the molecule, E_0 and ν_0 are the vibrational amplitude and frequency of incident light, q_0 and ν_m are the vibrational amplitude and frequency

of the molecule, respectively. The first term is named Rayleigh scattering, which occurs at the same frequency as the incident light, and has a magnitude proportional to α_0, the inherent polarizability of the molecule. The second term is anti-Stokes Raman scattering, which occurs at $\nu_0+\nu_m$ and the third term is Stokes Raman scattering at $\nu_0-\nu_m$.[57]

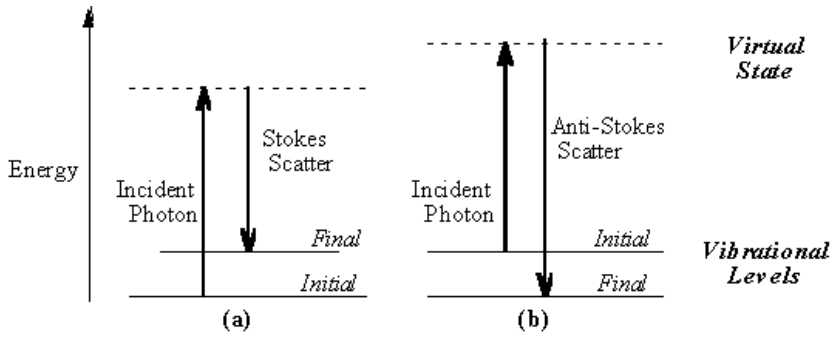


Figure 2.3 Energy transition diagram showing (a) Stokes and (b) anti-Stokes Raman scattering.

Figure 2.3 depicts the Jablonski diagram of conventional Raman scattering. Briefly, the incident photon excites the molecule to a short living virtual energy state. If the energy of the incident photon is unchanged after interaction with the molecule, Rayleigh scattering (i.e., elastic scattering) occurs. However, depending on the specific vibrational mode and molecular configuration a very small proportion of incident photons (~1 in 10^8) can be scattered with a change in frequency (i.e., Raman scattering). The absolute Raman scattering cross-section of molecules depends on the polarizability change associated with a vibrational mode. The simple classical description above (or semi-classical descriptions) however does not explain absolute intensities and why certain selection rules occurs as observed in experimental Raman spectroscopic data.[58] Hence, a rigorous description based entirely on quantum mechanics is necessary to fully appreciate experimental Raman data.[58] Kramer-Heisenberg dispersion (KHD) relation is the fundamental equation in scattering theory and can be derived analytically through time-dependent perturbation theory of the

interaction between a quantized electronic system (i.e., the molecular Hamiltonian) and the second quantized photon field (i.e., field Hamiltonian).[59] Briefly, the energy states of any molecule can be resolved into internally coupled electronic, vibrational and rotational energies. In adiabatic formalism (i.e., Born Oppenheimer approximation), the electronic state can be separated from the vibrational and rotational states.[59] The second quantization of the photon field (i.e., field Hamiltonian) can be represented by an infinite number of harmonic field oscillators with different frequencies.[59] The transition probability amplitude of Raman scattering can be determined using time dependent perturbation theory.[59] A rigorous derivation of KHD equation from first principles can be found elsewhere.[58, 59] Equation [3] shows the typical form of the KHD equation (i.e., differential scattering cross section) used in the literature in bra-ket notation.

$$\frac{I_{\text{incoming}}}{I_{\text{scattered}}d\Omega} = (e\omega_{k'})^4 \left| \sum_r \frac{\langle b|\vec{\hat{r}}_e \cdot \vec{u}^*_{k'}|r\rangle\langle r|\vec{\hat{r}}_e \cdot \vec{u}_k|a\rangle}{E_r - E_a - \hbar\omega_k - i\gamma} + \frac{\langle b|\vec{\hat{r}}_e \cdot \vec{u}_k|r\rangle\langle r|\vec{\hat{r}}_e \cdot \vec{u}^*_{k'}|a\rangle}{E_r - E_a + \hbar\omega_k + i\gamma} \right|^2 \qquad [3]$$

Where *a* is the initial state of the molecule, *b* is the final state of the molecule, *r* is the virtual state, ω_k is the angular frequency for the incoming photons, $\omega_{k'}$ is the angular frequency for the scattered photons. E_a, E_b and E_r are the energies of initial, final and virtual states respectively. γ is a dampening of the virtual state and $\vec{u}_k$ and $\vec{u}^*_{k'}$ are the polarization vectors for the incoming and scattered photon, $\vec{\hat{r}}_e$ is the electrical dipole operator. In brief, the Raman differential scattering cross-section is here given as the square of a coherent superposition of all molecular eigenstates. Therefore Raman scattering is quantum mechanically considered a "coherent absorption-emission" process. This equation is complex as the square of cross-terms appears. Important experimental effects can be derived from the equation: (i) in principle, all eigenstates of the molecule contributes in a coherent manner to the scattering process (ii) when

the incident light is in resonance with an electronic state, the scattering cross section is enhanced (i.e., resonance Raman effect). (iii) The Raman scattering cross section is inversely proportional to the fourth power of laser wavelength. The above quantum mechanical descriptions of the Raman scattering process allow one to rightly appreciate the high molecular specificity of the Raman scattering effect.

2.3.2 Raman instrumentation for tissue measurements

Raman spectroscopy requires monochromatic light to produce a well resolved spectrum.[60] Lasers produce a coherent beam of monochromatic light and are therefore used as an excitation source. For bulk structural investigations, visible excitation is typically utilized (e.g., Ar+ ion - 488 and 514.5 nm, Kr+ ion - 530.9 and 647.1 nm, He-Ne laser - 632.8 nm) for conventional Raman spectroscopy or resonance Raman spectroscopy. For biological application, NIR excitation is most commonly used (i.e., 785, 830 or 1064 nm). This is mainly due to the low light absorption of tissue in this spectral range (i.e., deeper tissue penetration due the tissue-optical window) as well as prominent reduction in interfering autofluorescence compared with visible excitation.[57] Nowadays, compact electronically controlled wavelength stabilized and narrow line-width NIR diode lasers have replaced the gas and neodymium-doped yttrium aluminum garnet (Nd:YAG) lasers because of their low cost, compact sizes and reliability.[60] The development of high-throughput imaging spectrographs based either on volume phase holographic (VPH) transmission or reflective type gold coated gratings with high efficiency has recently paved for rapid Raman spectroscopy techniques in medicine.[57] High-throughput imaging spectrographs with low f/# (e.g., f/1.8 or f/2.2), low stray light background (10^{-11}), NIR optimized optics are now being produced by several manufactures. Spectral range coverage varies according to the choice of grating (e.g., either fingerprint (FP) region (0 - 2000 cm^{-1}), high-

wavenumber (HW) region (2800 - 4000 cm^{-1}) or both). The spectrometer resolution is determined by several complex factors (including grating, entrance slit width and CCD pixel size).[57] CCD detection systems have generally replaced the single-channel detector (e.g., photomultiplier tubes (PMT) etc.). CCD's are most commonly characterized by the detection sensitivity (i.e., quantum efficiency (QE) at a certain temperature) and dark current rate.[61, 62] To reduce dark current, thermoelectric or cryogenic cooling of the CCD is commonly used. Since development of miniaturized fiber-optic probes for endoscopic sampling remains one of the major challenges in medicine, the following section will discuss fiber-optic Raman probes for endoscopic applications in more details.

2.3.3 Fiber-optic Raman probes for endoscopy

The difficulty in developing miniaturized flexible fiber-optic Raman probe designs with high collection efficiencies that are accessible in endoscopic applications have hindered the routine use of Raman spectroscopy in clinical settings.[63, 64] Comprehensive reviews of fiber-optic probes for biomedical spectroscopy can be found by Utzinger et al and Krafft et al.[63, 65] There are several important design considerations for fiber-optic Raman probes in endoscopic applications:[66-68] (i) The probe must have high collection efficiency to be able to acquire Raman spectra in very short acquisition time (<1-0.5 sec). (ii) Integration of distal band pass and long pass filters is common to reduce laser noise, Rayleigh scattered light as well as fused-silica Raman and fluorescence interference. (iii) There are strict geometry considerations for clinical use to fit into the instrument channel of endoscopes (e.g., endoscope, bronchoscope, colonoscope etc.). For instance, in gastroscopes, the instrument channel is ~2.2 mm (iv) The fiber-optic probe must be flexible, bio-

compatible and able to withstand medical sterilization. Zhang et al. were one of the first to test different designs of endoscopic-type fiber-optic probes (single fiber, simple bifurcated probe and bifurcated probe with narrow band coating at the excitation fiber tip) on chicken fat tissue.[67] The bifurcated volume-type fiber bundle with separate illumination and collection fibers were proposed to reduce the intrinsic fiber Raman and fluorescence signal. Zhang et al. found that the miniaturized bifurcated multi-fiber probe with filters in front of the excitation fiber and collection fibers could reduce the interference of fiber silica fluorescence and Raman signals.[67]

Since carcinogenesis typically initiates in the epithelial layer it has been realized that controlling the probing depth and sampling volume is clinically very important. Several design strategies such as beveled collection fibers, have been proposed to maximize shallower tissue Raman measurements in the epithelial tissue. In 1999, Wilson et al. developed a beveled endoscopic fiber probe[69] and performed the pioneering work on *in vivo* Raman endoscopic spectral measurements from the GI tract.[70] Wilson et al. used a 1.5 m long endoscopic compatible probe that consisted of a single 400 μm diameter central delivery fiber, surrounded by seven 300 μm diameter beveled collection fibers in order to realize high collection efficiency. Wilson and co-workers utilized the probe *in vivo* and successfully detected dysplastic lesions in BE.[71] Apart from the beveled collection fibers, the use of gradient index (GRIN) lenses and ball lens-coupled Raman probes have also been proposed as alternative methods to improve depth selectivity and collection efficiency. In 2004, Motz et al. presented a 2 mm diameter sapphire ball lens coupled fiber-optic probe[66] equipped with a rod filter module. The distal ball lens was utilized to improve the performance and collection efficiency of the probe. The diagnostic utility of the probe was tested through simulations, tissue model experiments and *ex vivo* tissues such as

artery and breast tissue samples.[66] Short et al. developed an endoscopic fiber-optic catheter Raman probe featuring a 200 μm central excitation fiber with the surrounding 27 collection fibers (100 μm diameter each) to characterize lung tissue *in vivo*.[72] Day et al. have very recently designed a miniaturized confocal endoscopic Raman probe consisting of a single illumination fiber and a single collection fiber utilizing a GRIN lens for optimizing the depth of focus ~150 μm to study early esophageal carcinogenesis.[68] Krafft and colleagues reported on the miniaturized (~0.375 mm diameter) fiber-optic probe based on multi-core single-mode fibers (MCSMF) with inscribed in-line fiber Bragg grating (FBG).[73] The applicability of the developed probe for clinically relevant samples was tested on porcine brain tissue.

In general, the laser excitation power density on tissue should be kept within a range that does not generate cytotoxity in cells. At the time of writing there exist no guidelines or standards on the use of NIR lasers (e.g., 785 nm) in internal epithelial organs. For this reason, as a general guideline, one should not apply more than a maximum 1.5 W/cm^2 on the tissue sample which is the maximum permissible skin exposure limit set out by the American National Standards Institute.[74]

2.4 Multivariate analysis for tissue spectroscopic diagnosis

2.4.1 Spectral preprocessing

Since tissue Raman spectra are complex and represent a superposition of autofluorescence and weak Raman signals from a myriad of biomolecules it is important to apply sophisticated signal processing to standardize the Raman signal

prior to analysis. Preprocessing techniques serve to eliminate redundant spectral variation or noise in the raw Raman spectra. A multitude of preprocessing techniques exists that can be applied to datasets of Raman spectra such as noise removal techniques (e.g., Fourier filtering, median filtering, Savitzky-Golay filtering etc.), normalization (e.g., peak or area normalization, multiple scatter correction (MSC), extended multiple scatter correction (EMSC)).[75, 76] In particular, the developments of robust autofluorescence baseline removal techniques remain important in Raman spectroscopy.[64, 77] Tissue autofluorescence signals are typically many orders in magnitude more intense that the Raman peaks.[78] Removal of autofluorescence can be realized using different methodologies such as wavelet transformation,[79] shifted subtracted Raman spectroscopy (SSRS),[80] polynomial fitting,[64] Fourier transform, etc. Many research groups have realized that 4-6th order polynomial constrained to the lower portion of the calibrated spectra is a robust methodology to efficiently remove the autofluorescence signal in tissue Raman spectra in the FP region (e.g., 800-1800 cm^{-1}).[64] To demonstrate this, Figure 2.4 shows the 5th order fitting of the autofluorescence signal from GI tissue *in vivo.* The Raman spectrum is also shown after autofluorescence subtraction revealing the distinct Raman peaks with essentially no slope or baseline offset. Notice that although it is common to subtract/remove the tissue autofluorescence in the literature,[26-31] autofluorescence carries molecular information that in some applications could be complementary to the Raman signal.[78]

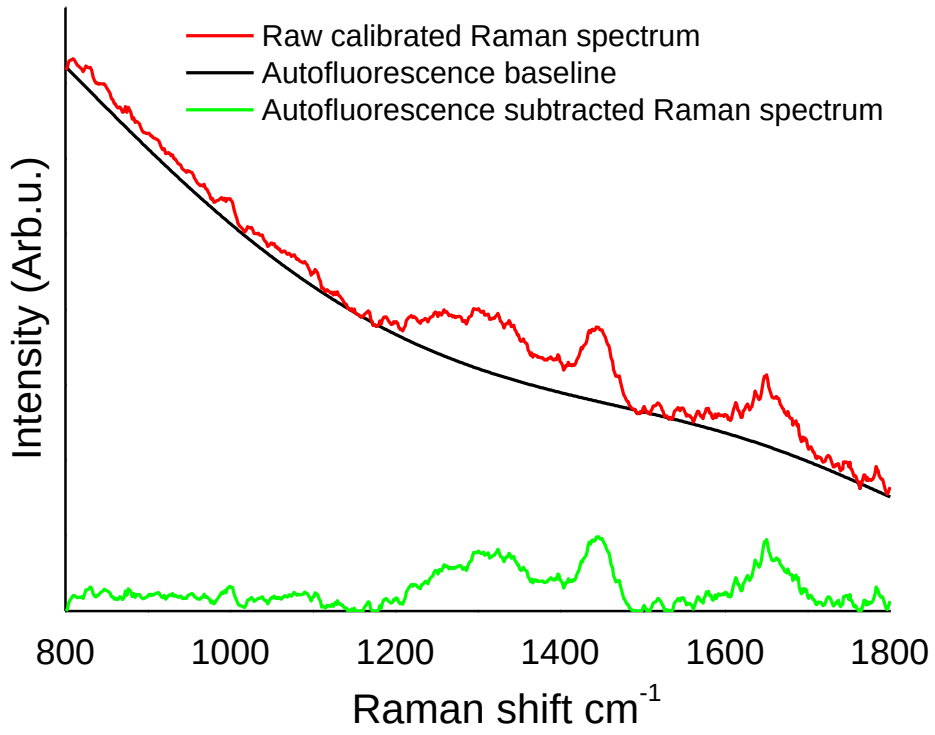


Figure 2.4 Autofluorescence background and Raman signal from gastric tissue. Raman signal after autofluorescence 5th order polynomial background subtraction iteratively constrained to the lower portion of the spectra.

2.4.2 Principal component analysis (PCA)

Empirical approaches to analyze Raman spectra (e.g., intensity peaks or intensity ratio of Raman peaks) have been used to classify Raman spectra according to tissue pathology types.[27] Simple intensity ratio algorithms provide tissue information which is independent of Raman spectroscopic measurement conditions (e.g., absolute intensities).[26] Multivariate/chemometric analysis is however needed to extract biomolecular information for clinical diagnostics. Principal component analysis (PCA) has become a standard technique for multivariate/chemometric analysis of optical spectroscopic data.[81] Briefly, this mathematical technique reduces the dimension of the entire Raman spectrum to a reduced number of data points by finding eigenvectors and eigenvalues to the covariance matrix of the dataset.[82] That is,

extracting the largest spectral variations successively. Accordingly, PCA decomposes the spectra into loading and scores (i.e., principal components (PCs)), in matrix notation:

$$\mathbf{X} = \mathbf{TP}^{\mathbf{T}} + \mathbf{E} \quad [4]$$

where **T** and **P** represent scores and loadings, and **E** contains the residuals. For instance, loading one (i.e., the first vector in **P**) represents the largest spectral variation in the data and the score value is the projection of the spectrum on the loading. This applies successively for each PC. In other words, the loadings correspond to a new rotated coordinate system, and scores represent the data projection values. Choosing the number of PCs for diagnostic purposes remains tentative and must be based on common sense to avoid overfitting the data. Most literature studies use analysis of variance (ANOVA) and *p*-values among different groups to choose the PCs.[28] PCA can be combined with clustering or linear/non-linear discriminant analysis for classification.[81] The PCA scores can in principle be fed to any clustering analysis (e.g., K-nearest neighbours, etc.). Other useful classification techniques include for instance support vector machines (SVM), which maps the spectra into a higher dimensional space through either a linear or nonlinear kernel,[83] or classification and regression trees (CART) as well as swarm intelligence or genetic algorithms (GA) for variable/feature selection.[30, 31, 84-89]

2.4.3 Partial least squares discriminant analysis (PLS-DA)

Partial least squares (PLS) is a powerful regression method for spectral data. PLS can also advantageously be applied for discriminant analysis (DA) problems by encoding the class membership of zeros and ones, representing group affinities in an

appropriate **Y**-indicator matrix. PLS-DA employs the fundamental principle of PCA but further rotates the coordinate system (i.e., latent variables (LVs)) by maximizing the covariance between the spectral variation and group affinity so that the LVs explain the diagnostic relevant variations rather than the most prominent variations in the spectral dataset.[90] The general model is given in matrix notation as equation [5]

$$\mathbf{X} = \mathbf{TP}^{\mathbf{T}} + \mathbf{E}$$
$$\mathbf{Y} = \mathbf{UQ}^{\mathbf{T}} + \mathbf{F} \qquad [5]$$

where **T** and **P** represent scores and loadings for the spectral data, and **E** contains the residuals. **U** and Q represent scores and loadings for the affinity matrix, and **F** contains the residuals. This model allows diagnostic information to be efficiently compressed in the first few LV's. The dichotomous PLS-DA algorithm can be extended to multi-class classification using one-against-one or one-against-all scheme with max-wins voting or other voting strategies.[91] Many useful extensions of PLS for optimization exist including Interval PLS-DA, PLS-GA etc.[84]

2.4.4 Diagnostic performance validation

Independent validation is necessary to obtain accurate estimates of diagnostic algorithm performance. This is however not always possible due to practical considerations (e.g., limited data size, economic reasons etc.). Hence to estimate the performance (i.e., accuracy, sensitivity and specificity) of algorithms (and determine optimal algorithm complexity for PCA or PLS-DA etc.), leave-one-out cross-validation can be employed.[82] In this technique, spectra from one patient (or site) are left out and the model is rebuilt with the remaining spectra. The left-out spectra are then predicted and this is repeated iteratively through all patients/sites to obtain an estimate of diagnostic algorithm performance. Independent block validation can also

be performed by splitting the dataset into training (e.g., 80% of dataset) and testing (e.g., 20% of dataset) and this can be repeated iteratively to obtain confidence in the estimates.[43] In this dissertation, we also employ receiver operating characteristics curves (ROC).[92] ROC analysis offers efficient tools to select optimal classification models. ROC curves are calculated by changing the classification thresholds and plotting the sensitivity as a function of 1-specificity. The area under the curve (AUC) can be calculated to assess the robustness of diagnostic models.[92]

2.4.5 Biomolecular and spectral modeling

Biological Raman spectra are intrinsically complex to interpret due to a multitude of overlapping peaks mainly of protein, lipids and DNA.[25] However, to evaluate gross quantitative differences between different biological tissue types, semi-quantitative spectral modelling of the tissue Raman spectra using basis reference biochemicals can be employed.[93] The non-negativity least squares minimization (NNLSM) modelling can be used to grossly fit representative biochemicals to the tissue Raman spectra (Equation [6])

$$\mathbf{X} = \mathbf{cS} + \mathbf{E} \text{ where } \mathbf{c} \geq 0, \quad [6]$$

Where **c** is the fit-coefficients, **S** are the basis spectra, **E** are the residuals and **X** are the tissue Raman spectra, The resulting fit-coefficients are constrained to a total sum of 100% so that relative variations in tissue can be evaluated. One should note that biochemical modelling of tissue spectra by means of NNLSM is a linear semi-quantitative estimation of biochemical compositions. Only the most essential biochemical constituents of tissue (i.e., biochemicals largely representing cell nucleus, cell cytoplasm, and intra-/extra-cellular matrices) should be included in the NNLSM

modeling. The discrepancy between the reconstructed Raman spectra (i.e., by the NNLSM modeling) and the measured Raman spectra may be due to the following reasons: (i) The *ex vivo* biomolecular compositions and conformations used for NNLSM modeling may not truly reflect *in vivo* tissue conditions; (ii) The basis spectra are not orthogonal, and an incomplete number of biochemicals are included; (iii) The NNLSM biochemical modeling does not take the effect of tissue optical properties (e.g., absorption and scattering coefficients, and anisotropy) of different tissues into consideration; (iv) The laser penetration depth may also affect the estimated distributions of the biochemicals in different tissue types. Nevertheless, this semi-quantitative spectral analysis can give a fair estimate of the overall protein, lipid and DNA contents in tissue.[93, 94]

2.5 Overview of *in vivo* Raman endoscopy literature studies

Accumulating evidence has confirmed the clinical potential of Raman spectroscopy *ex vivo*. A large number of *ex vivo* Raman spectroscopy studies on excised tissues and cells have been reported on the GI system.[26-31] For instance, Stone and co-workers have characterized esophageal cells and tissue specimens associated with Barrett's disease.[37, 94-97] Raman microspectroscopy was used to map the biomolecular distribution of protein, lipid and DNA in sectioned esophageal tissue samples, revealing the distinct biomolecular features associated with Barrett's carcinogenesis.[94] Our group has reported a series of *ex vivo* tissue Raman studies on gastric tissues.[26-31] For instance, Teh et al., reported the differentiation among different types of pathologic gastric tissues (e.g. HP infection, IM, dysplasia, neoplasia as well as differentiation between diffuse and intestinal-type adenocarcinoma) with good accuracies (of ~85–96%).[26-31] Raman spectroscopy has also shown promising results

for differential diagnosis between adenomatous and hyperplastic polyps in the colon *ex vivo*.[83, 98] Wilson and co-workers[69, 70, 99-102] reported a Raman study on eight patients, showing that Raman spectroscopy could distinguish adenomatous from hyperplastic polyps *ex vivo* with 91% sensitivity and 95% specificity. In general *ex vivo* studies have been used to evaluate the diagnostic capability and characterize different tissue types using the highly specific Raman peaks. Since the Raman peaks of tissue and cells (mainly of protein, DNA and lipids) are superimposed, assignments of biochemicals/biomolecules and vibrational modes are tentative in nature. Table 1 summarizes the location of the most prominent tissue Raman peaks in the spectral range (400-3600 cm^{-1}) reported from the literature (accuracy of ~1-3 cm^{-1}) with tentative vibrational/biomolecular assignments. The position of these specific Raman peaks listed can serve as a foundation for characterization of *ex vivo* and *in vivo* tissue Raman spectra.

Peaks (~cm^{-1})	**Tentative vibrational/biomolecular assignments**	**References numbers**
429	Cholesterol	[103]
490	Glycogen	[96]
509	υ(S–S) of proteins	[104]
669	υ(C–S) of cystine	[96]
752	Ring breathing mode of tryptophan	[25]
782	υ(O–P–O) of DNA	[105]
853	υ(C–C), proline	[25]
875	υ(C–C), hydroxyproline of proteins	[27, 28]
936	υ(C–C), α-helix, proline, valine	[25]
956	υ_s(P=O) of hydroxyapatite, bone	[106, 107]
1004	υ_s(C–C), symmetric ring breathing, phenylalanine	[27, 28]

1078	υ(C–C) of lipids	25
1085	υ(C–N) of proteins	96, 97
1123	Glucose	108
1245	υ(C–N), δ(N–H) Amide III	27
1302	$\delta(CH_2)$ deformation of proteins and lipids	97
1335	CH_3CH_2 wagging mode of proteins and nucleic acids (DNA-purine bases)	27, 109
1370	Adenine, thymine, guanine (DNA, RNA)	97, 109
1420	Adenine, guanine (DNA, RNA)	110
1445	$\delta(CH_2)$ deformations of proteins and lipids	27, 96, 97
1558	Tryptophan	111
1579	Guanine of DNA	96, 109
1618	υ(C=C) of porphyrins (hemoglobin)	28
1638	υ_2 bending mode of water	112
1655	υ(C=O) Amide I of proteins	25, 27, 93
1745	υ(C=O) of phospholipids	27, 93
2850	$\upsilon_s(CH_2)$ of lipids	113-115
2885	$\upsilon_{as}(CH_2)$ of lipids	113, 116
2939	$\upsilon(CH_2)$ of proteins	113
3011	υ(=C–H) of lipids	113, 114
3232	υ(O–H) and υ(N–H) vibrations of water and proteins	113
3400	υ(O-H) vibrations of water	113, 116, 117

Table 1 Tentative vibrational/biomolecular peak assignments of tissue Raman spectra in the range 400-3600 cm^{-1}

While tissue characterization and peak assignments remain important fields in biomedical Raman spectroscopy, the ultimate goal is clinical *in vivo* implementation. To move the Raman technique into medical diagnosis at endoscopy, Shim et al.

pioneered the *in vivo* fiber-optic Raman spectroscopy technique for GI endoscopic applications,[70, 102] and demonstrated the capability of Raman spectroscopy for the differential diagnosis between adenomatous and hyperplastic polyps in the colon *in vivo* with a sensitivity of 100% and specificity of 89%.[118] However, Raman acquisition times in their Raman system design was lengthy (>5 sec), which was impractical for routine real-time endoscopic applications.[64] Hence, additional technological development was needed to transfer Raman spectroscopy into clinical endoscopy. With very recent technological advancements in Raman spectroscopy instrumentation, including high-throughput spectrographs, NIR lasers, sensitive CCD cameras, personal computers and fiber-optic Raman probes, significant progress has been made in translating Raman spectroscopy into real-time *in vivo* routine endoscopic applications. Accumulating evidence from different organs has now confirmed the clinical potential of Raman spectroscopy *in vivo*.[43, 50, 78, 89-91, 106, 112, 119-125] Table 2 summarizes the *in vivo* studies and most important experimental parameters and results reported from different endoscopic accessible organs (i.e., esophagus, gastric colon, lung and bladder) identified using a comprehensive PubMed search.

Organ	Patients recruited, spectra	Experimental parameters	Diagnostic outcomes	Year, reference
Larynx	39 patients 94 spectra	Power density: 1.5W/cm^2 Integration time: <1 s	Identification of laryngeal cancer with a sensitivity of 90.3% and a specificity of 90.9% using PCA-LDA	2012 [115]
Esophagus	65 patients 192 spectra	Power: 70mW Integration time: 5 s	Dysplastic BE could be differentiated from non-dysplastic BE with a sensitivity of 86% and a specificity of 88% using GA-ORS/LDA	2005 [71]
	107 patients 263 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Sensitivity of 90.9% and a specificity of 93.9% using PLS-DA for cancer detection	2011 [121]
	27 patients 75 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Sensitivity of 97.0% and a specificity of 95.2% for cancer detection using NNLSM-LDA	2011 [122]
	77 patients 797 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Prospective diagnostic sensitivity of 87.0% and specificity of 84.7% can be obtained for *in vivo* detection of HGD in BE using the PLS-DA model developed from 373 patients	2014 [126]
Gastric	30 patients 72 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Sensitivity of 94.4% and a specificity of 96.3% for detecting dysplasia using PCA-LDA	2010 [123]
	71 patients 1102 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Sensitivities of 90.8%, 84.7%, 82.1%, and specificities of 93.8%, 94.5%, 95.3% for classification of normal, benign and malignant gastric ulcers respectively, using PLS-DA	2010 [90]
	107 patients 926 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Sensitivity of 92.6% and a specificity of 88.6% for gastric cancer detection using PLS-DA	2011 [121]
	81 patients 1238 spectra	Powerdensity:1.5 W/cm^2 Integration time: <0.5 s	Sensitivity of 97.9% and a specificity of 91.5% for cancer detection using PCA-LDA	2011 [78]
	10 patients 75 spectra	Power density:1.5W/cm^2 Integration time: <0.5 s	Prospective sensitivity of 90.0% and specificity of 73.3% based on the generated PLS-DA model from the spectral library (2748 spectra measured from 305 patients) for cancer detection.	2012 [43]

	83 patients 1277 spectra	Power density:1.5 W/cm^2 Integration time: <0.5 s	PLS-DA algorithm detects IM, dysplasia and cancer with sensitivities of 46.67%, 83.33%, 84.91% and specificities of 87.55%, 95.80%, 95.57%, respectively.	2013 [91]
Colon	3 patients 19 spectra	Excitation power: 100 mW Integration time: 5 s	Sensitivity of 100% and specificity of 89% for discriminating adenoma from hyperplastic polyps	2003 [118]
Lung	46 patients 129 spectra	Maximum excitation power: 200 mW Integration time: 1 s	Sensitivity of 96% and a specificity of 91% using PCA-LDA for detecting preneoplastic lesions	2011 [127]
Bladder	38 patients 63 spectra	Excitation power: 56-89 mW Integration time: 1-5 s	Sensitivity of 85% and specificity of 79% for detecting bladder cancer using PCA-LDA	2010 [128]

Table 2 Summary of reported *in vivo* Raman spectroscopic studies for the diagnosis of endoscopic precancers and cancers to date. Important experimental parameters and results are given.

Besides in the GI tract, Raman spectroscopy has shown promising potential in other endoscopic accessible organs. For instance, Zeng and co-workers have developed an image-guided Raman endoscopic system for bronchoscopic applications.[129] They conducted a study based on 46 patients, and reported that Raman endoscopy complements WLR and AFI by reducing the false positives biopsies.[127] WLR+AFI and Raman endoscopy showed sensitivity and specificity of 96% and 91% respectively for detecting the preneoplastic lesions.[127] Their study was however not randomized. Draga et al. conducted an *in vivo* Raman study on 38 patients in the bladder during the procedure of transurethral resection of bladder tumors (TURBT), and found that Raman spectroscopy could differentiate bladder cancer from normal tissue with a sensitivity of 85% and a specificity of 79%.[128] Very recently, our group has reported on the implementation of transnasal Raman endoscopy of nasopharyngeal and laryngeal tissues.[112] Our group conducted a study on 39 patients and measured a total of 94 Raman spectra.[115] Raman spectra from tumors showed

distinct spectral characteristics, and could be diagnosed on spectrum basis with a sensitivity of 90.3% and a specificity of 90.9%. In general Raman spectra are highly specific and contain a wealth of biomolecular diagnostic information.

Chapter 3. Outline of clinical study protocols

The present study was conducted in close collaboration with the Endoscopy Centre at the National University Health System (NUHS), Singapore, and is part of a continuous nationwide study focusing on early diagnosis and treatment of gastrointestinal malignancies run by the Singapore Gastric Cancer Epidemiology, Clinical and Genetic Program (GCEP). GCEP aims to identify an optimum approach and cost-effective measures for targeted screening of gastric cancer in the local population. The GCEP Cohort Study for subjects at high risk of gastric cancer was initiated in January 2004, with a planned enrollment of 4000 well-characterized high-risk patients undergoing systematic screening for early gastric cancer, supported by an annotated clinical database and the banking of high-quality gastric tissue and plasma specimens. The GCEP Cohort Study enrolls people aged >50 years who are at high risk of gastric cancer and offers screening by endoscopy with systematic prospective follow up over a minimum of 5 years. Participation is voluntary and participants receive detailed explanation and counseling on the benefits of the study, as well as the requirements of the study. This study has been approved by the Institutional Ethics Review Boards (IRB) in four major hospitals in Singapore. The study captures comprehensive clinical information on a state-of-the-art database, and obtains prospectively blood, serum, white-blood-cell DNA, gastric biopsy tissue, and gastric-juice samples according to a standardized protocol.

This Raman work was approved by the Institutional Review Board (IRB) of the National Healthcare Group (NHG) of Singapore. The trial was performed in accordance with International Conference on Harmonization (ICH) for Good Clinical Practice (GCP) guidelines, Declaration of Helsinki (2000). All participants were

orally informed and have signed a written informed consent. The study protocols, written informed consents, and record holdings are stored by the Gastrotranslational Office, NUHS. Raman spectral data are stored in a database in the Department of Biomedical Engineering, Bioimaging Laboratory, National University of Singapore (NUS). A supplementary (SR) protocol is used to recruit in-patients at the Endoscopy Center, NUHS. From August 2008 to June 2013, a total of 450 patients have been enrolled in the Raman endoscopic examinations at NUHS. We have systematically conducted Raman measurements of both the esophagus and gastric from these patients. Over 12,000 Raman spectra have been measured. The histopathological examinations from Department of Pathology, NUHS served as the gold standard for evaluating the diagnostic sensitivity and specificity of *in vivo* Raman endoscopy technique. The data presented in this dissertation (Chapter 4, Chapter 5, Chapter 6 and Chapter 7) has been analyzed continuously when accumulating *in vivo* Raman spectra and evidence have become available.

Chapter 4. Characterizing variability in *in vivo* Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection

This chapter aimed to develop the fiber-optic Raman endoscopy technique and characterize *in vivo* Raman spectroscopic properties of tissues belonging to different anatomical regions in the upper GI tract and explore the implications for early detection of neoplastic lesions during clinical gastroscopy. A novel fiber-optic Raman endoscopy technique was developed and applied for real-time *in vivo* tissue Raman measurements of normal esophageal (distal, middle, and proximal), gastric (antrum, body, and cardia) as well as cancerous esophagous and gastric tissues from 107 patients who underwent endoscopic examinations. The NNLSM coupled with a reference database of Raman active biochemicals (i.e., actin, histones, collagen, DNA, and triolein) was employed for semiquantitative biomolecular modeling of tissue constituents in the upper GI.

A total of 1189 *in vivo* Raman spectra were analyzed from different locations in the upper GI. The Raman spectra among the distal, middle, and proximal sites of the esophagus showed no significant interanatomical variability. The interanatomical variability of Raman spectra among normal gastric tissue (antrum, body, and cardia) was subtle compared to cancerous tissue transformation, whereas biomolecular modeling revealed significant differences between the two organs, particularly in the GE junction. Cancerous tissues could be identified across interanatomical regions with accuracies of 89.3% (sensitivity of 92.6% (162/175); specificity of 88.6% (665/751)), and of 94.7% (sensitivity of 90.9% (30/33); specificity of 93.9% (216/230)) in the gastric and esophagus, respectively, using PLS-DA together with the

leave-one tissue site-out, cross validation. This chapter demonstrates that Raman endoscopy technique has promising clinical potential for real-time, *in vivo* diagnosis and detection of malignancies in the upper GI at the molecular level.

4.1 Introduction

We present an integrated Raman spectroscopy and trimodal (WLR, NBI, AFI) imaging technique for real-time *in vivo* tissue Raman measurements at endoscopy. A special 1.8 mm endoscopic Raman probe with filtering modules is developed, permitting effective elimination of interference of fluorescence background and silica Raman in fibers while maximizing tissue Raman collections. High-quality *in vivo* Raman spectra of upper gastrointestinal tract can be acquired within 1 s or subseconds under the guidance of wide-field endoscopic imaging modalities, greatly facilitating the adoption of Raman spectroscopy into clinical research and practice during routine endoscopic inspections. The histological characteristics and morphologies of distinctive anatomical regions in the upper GI tract (i.e., esophagus and gastric) are highly functionally specialized and exhibit significant variations in architectural properties and cell types (e.g., tissue thickness variability, distinct glandular types, secretion products, vascularity, etc.)[130, 131]. Hence, to assess the potential of Raman endoscopy technique, the extent of *in vivo* Raman spectral interorgan and interanatomical variabilities of tissues in the upper GI tract must be validated. Hence, the main aim of this study was to evaluate the magnitude of interorgan and interanatomical variability of *in vivo* normal tissue Raman endoscopic spectra in the esophagus and gastric as well as to assess the implication for early diagnosis of neoplastic lesions. The semiquantitative biomolecular modeling NNLSM[132] using representative biochemical basis spectra is also employed for estimation of the most prominent Raman active constituents of tissue in the upper GI tract.

4.2 Materials and methods

4.2.1 Instrumentation

The novel integrated Raman spectroscopy and trimodal wide-field imaging system developed for *in vivo* tissue measurements and characterization at endoscopy is shown in Figure 4.1 and consists of a wavelength stabilized 785 nm diode laser (maximum output: 300 mW, B&W TEK Inc., Newark, DE), a transmissive imaging spectrograph (Holospec f/1.8, Kaiser Optical Systems) equipped with a liquid-nitrogen cooled, NIR-optimized, back-illuminated, and deep depletion charge-coupled device (CCD) camera (1340×400 pixels at 20×20 μm per pixel; Spec-10: 400BR/LN, Princeton Instruments), and a specially designed 1.8 mm Raman endoscopic probe for both laser light delivery and *in vivo* tissue Raman signal collection.

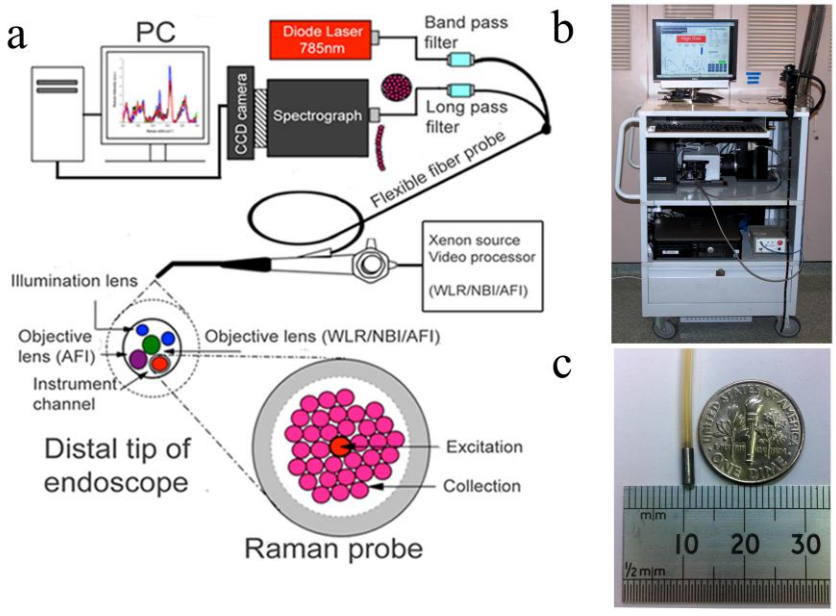


Figure 4.1 (a) Schematic of the integrated Raman spectroscopy and trimodal endoscopic imaging system developed for *in vivo* tissue Raman measurements at endoscopy; (b) Photo of the Raman endoscopic system in clinic; (c) Photo of the fiber-optic Raman endoscopic probe.

The novel Raman probe is composed of 32 collection fibers (200 μm, NA: 0.22)

surrounding the central light delivery fiber (200 μm, NA: 0.22) with two stages of optical filtering incorporated at the proximal and distal ends of the probe for maximizing the collection of tissue Raman signals while reducing the interference of Rayleigh scattered light, fiber fluorescence, and silica Raman signals. The Raman probe can pass down through the instrument channel of most medical endoscopes and be directed to all anatomical tissue sites under the guidance of wide-field endoscopic imaging (WLR/AFI/NBI) modalities[64] and the probe positioning against the mucosa can be verified on the endoscopy monitor by the clinicians in-charge during endoscopic examinations. Control of the *in vivo* Raman endoscopy system was implemented by a personal computer (PC) using custom-designed software that triggers on-line data acquisition and analysis (e.g., CCD dark-noise subtraction, wavelength calibration, system spectral response calibration, signal saturation detection, cosmic ray rejection, etc.), as well as real-time display of *in vivo* tissue Raman spectra during clinical endoscopic measurements.

The atomic emission lines from a mercury-argon spectral lamp (HG-1, Ocean Optics, Inc., Dunedin, FL) are used for wavelength calibration of tissue Raman spectra, and all wavelength-calibrated spectra were corrected for the wavelength-dependence of the system using a tungsten-halogen calibration lamp (RS-10, EG&G Gamma Scientific, San Diego, CA). The spectral resolution of the Raman system is ~9 cm^{-1} and the system acquires Raman spectra in the wavenumber range of 800–1800 cm^{-1} from *in vivo* GI tissue within 0.5 s using the 785 nm excitation power of 1.5 W/cm^2, which is less than the maximum permissible skin exposure limit set out by the American National Standards Institute.[74] This presumes the presence of melanin. Our further calculations based on the finite difference thermal model and the optical properties of the GI tissue[133, 134] indicate that even without consideration of other

cooling effects (e.g., perfusion and evaporation in tissue), the maximum tissue temperature rise is only about 0.15 ˚C after 1 min of 785 nm laser radiation with an incident power of 30 mW on a tissue spot size of 200 μm during tissue Raman measurements. This temperature rise estimated is far below the level to generate cytotoxicity in cells and tissue,[74] suggesting that the laser power density used in this study is safe for *in vivo* tissue Raman measurements in the upper GI.

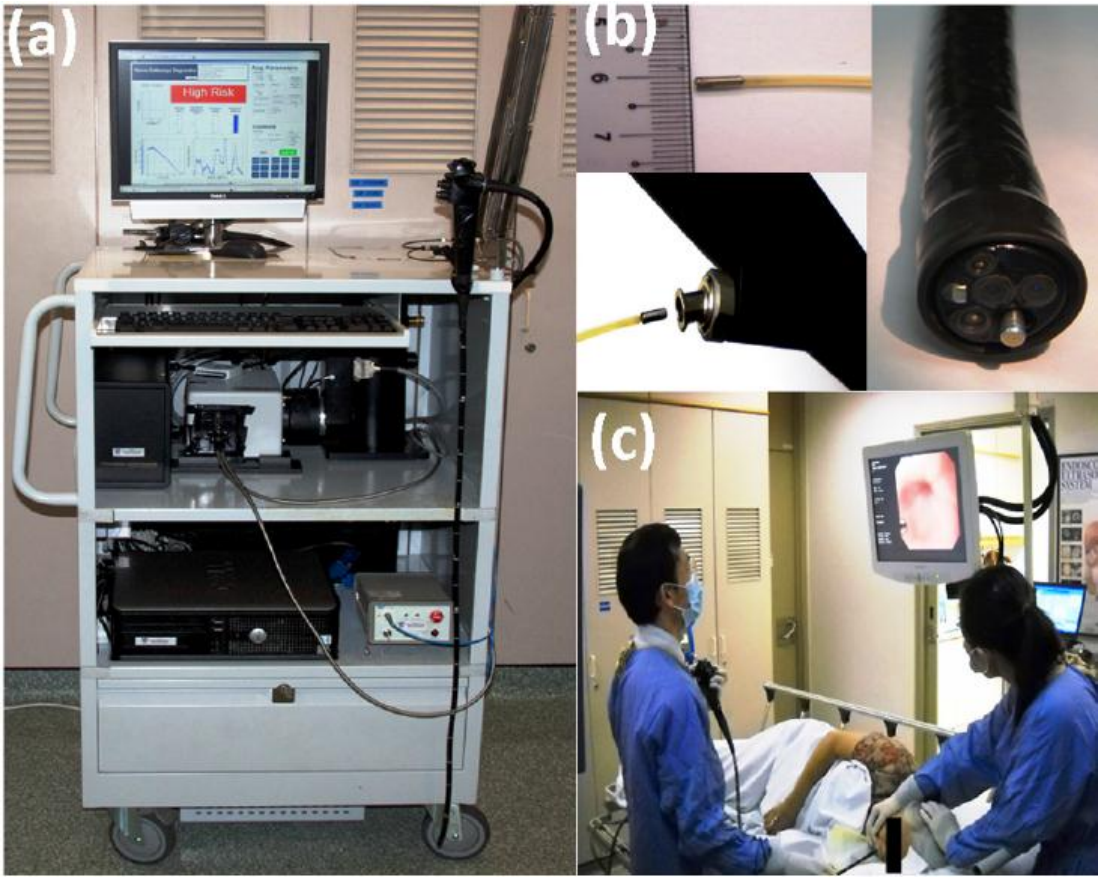


Figure 4.2 (a) Photograph of Raman endoscopy system in clinic; (b) insertion of the 1.8 mm Raman endoscopic probe into the working channel of an endoscope during gastroscopy; and (c) routine Raman endoscopy procedure in clinic.

The trimodal endoscopy imaging system comprises a 300 W short-arc xenon light source, a gastrointestinal videoscope (GIF-FQ260Z, Olympus), and a video system processor (CV-260SL, Olympus). The light reflected or autofluorescence emitted from tissue are detected by two monochrome CCD chips mounted behind the

two objective lens placed next to each other at the distal tip of the GI videoscope: one CCD for WLR/NBI and the other one for AFI. With this unique image-guided Raman endoscopy system, wide-field endoscopic images (WLR/AFI/NBI) and the corresponding *in vivo* Raman spectra of the tissue imaged can be simultaneously acquired, displayed, and recorded in the video system processor and the PC, respectively. We have also integrated the system on a cart (Figure 4.2) with an axillary power unit (APU) that makes the Raman spectroscopy system movable in clinical settings.

4.2.2 Preprocessing

The raw Raman spectra (800–1800 cm^{-1}) measured from *in vivo* GI tissue represents a composition of weak Raman signals, intense autofluorescence background and noise.[64] Thus, the raw spectra were preprocessed by a first-order Savitsky-Golay smoothing filter (window width of five pixels, which corresponded to the system spectral resolution) to reduce noise.[64] A fifth-order polynomial was found to be optimal for fitting the autofluorescence background in the noise-smoothed spectrum,[64] and this polynomial was then subtracted from the raw spectrum to yield the tissue Raman spectrum alone. Each background-subtracted Raman spectrum was also normalized to the integrated area under the curve from 800–1800 cm^{-1}, enabling a better comparison of the spectral shapes and relative Raman band intensities among different organs and anatomical regions.

4.2.3 Study protocol

Outline of the clinical study protocol used can be found in Chapter 3. In this study, spectra from 107 patients (61 men and 46 women with a median age of 66 years old) were analyzed.[64] A total of 1189 *in vivo* Raman spectra were acquired and

analyzed from the upper GI tract including normal (antrum sites with greater and lesser curvatures, body sites with greater and lesser curvatures, and cardia) and cancerous gastric tissues, as well as normal (distal end, middle sites, and proximal sites) and cancerous esophageal tissues. Table 3 summarizes the number of *in vivo* Raman spectra acquired and the corresponding integration time required for each tissue type in the upper GI.

Tissue type	Number of Raman spectra acquired	Integration time (s)	Patient numbers
	Gastric		
Antrum (normal)	377	0.5	62
Body (normal)	268	0.2	42
Cardia (normal)	106	0.3	26
Cancer	175	0.5	35
	Esophagus		
Distal (normal)	57	0.4	13
Middle (normal)	66	0.4	15
Proximal (normal)	107	0.4	19
Cancer	33	0.5	15

Table 3 The number of *in vivo* Raman spectra acquired and the corresponding mean integration time required for each tissue type in the gastric and esophagus.

Immediately after Raman endoscopic acquisitions, the biopsies were taken from the tissue sites measured (with suction markings) and fixed in 10% formalin solution for histopathological confirmation of normal or malignant tissue types by a senior gastrointestinal pathologist who was blinded to the Raman spectroscopy results. Note that the cancerous gastric tissues (adenocarcinomas (n = 175)) were found in different sites (e.g., antrum/body/cardia) of the stomach; whereas cancerous esophageal tissues

(including adenocarcinomas (n = 26) and squamous cell carcinoma (n = 6)) were localized in the distal and middle parts of the esophagus, respectively. The histopathological examinations served as the gold standard for evaluating the diagnostic sensitivity and specificity of *in vivo* Raman endoscopy technique.

4.2.4 Biomolecular modeling

The biomolecular characterization of esophageal and gastric tissue is essential for implementation of the Raman endoscopy technique as a clinical tool for *in vivo* diagnosis of cancer and precancer in the upper GI. The biochemical spectral fitting by means of Raman spectra of cell constituents (e.g., cytoplasm, nucleus, etc.) and pure biochemicals has been extensively employed to gain better understanding of distinctive tissue constituents associated with pathologic changes.[94, 122, 132] In this chapter, we employ a semiquantitative model of *in vivo* tissue Raman spectra based on *a priori* insight of inter-/intra-cellular constituents using a linear combination (i.e., NNLSM)[132] of basis Raman spectra that represent the main biochemical constituents in the GI tract. Of over 35 basis reference Raman spectra obtained from different biomolecules associated with GI tissue (e.g., actin, albumin, pepsin, pepsinogen, nicotinamide adenine dinucleotide (NADH), RNA, DNA, myosin, hemoglobin, collagen I, collagen II, collagen V, mucin 1, mucin 2, mucin 3, flavins, elastin, phosphatidylcholine, cholesterol, glucose, glycogen, triolein, histones, beta-carotene, etc.), our NNLSM modeling indicate that the following five biochemicals, i.e., actin (A3653), histones (H6005), collagen type I (C9879), DNA (P4522), and triolein (T7140) (Sigma, St Louis, MO), were the most significant Raman-active biochemical constituents (Figure 4.3) that can effectively characterize gastric and esophageal (normal and cancerous) tissue with very small fit-residuals. For instance, DNA

represents nucleic acids within the cell nucleus; triolein represents typical lipid signals; actin and histones resembles proteins of different conformations and are the major components of the cytoskeleton and chromatin, respectively, whereas collagen type I is a substantial part of the extracellular matrix. Note that the basis Raman spectra of these biochemicals were measured in their native conditions without any further purification using the Raman endoscopy instrumentation previously described for semiquantitative biochemical modeling. The five coefficients rendered using NNLSM were further constrained to a sum of 100% to assess relative changes.

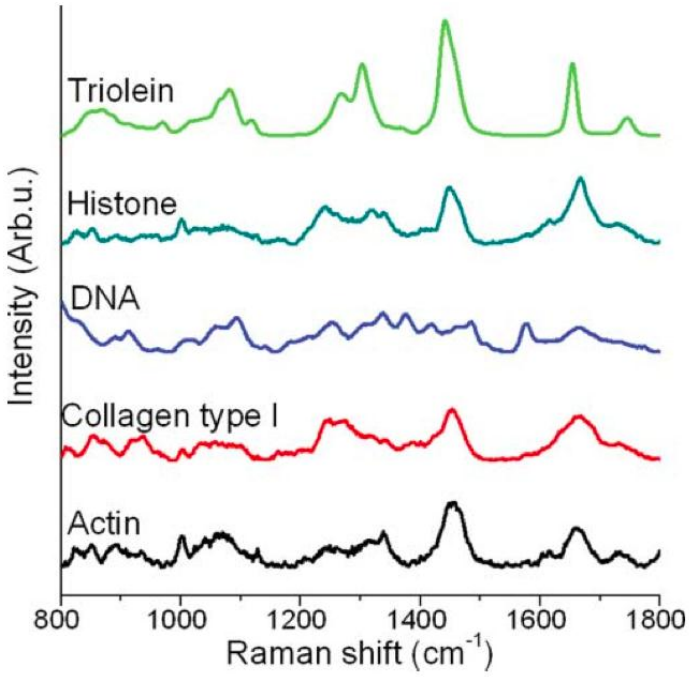


Figure 4.3 The five basis reference Raman spectra (i.e., actin, collagen, DNA, histones, and triolein) are used for biochemical modeling of esophageal and gastric tissue.

4.3 Results

High quality *in vivo* Raman spectra can routinely be acquired from different anatomical locations or lesion sites within the esophagus and gastric in real-time (<0.5 s) under multimodal imaging (WLR, NBI, and AFI) guidance during clinical

endoscopic examinations. Figure 4.4a shows the *in vivo* mean Raman spectra ± 1 standard deviations (SD) of normal (antrum, body, and cardia) and cancer gastric tissues; whereas Figure 4.4b shows the *in vivo* mean Raman spectra of cancer and normal (distal, middle, and proximal) esophageal tissues.

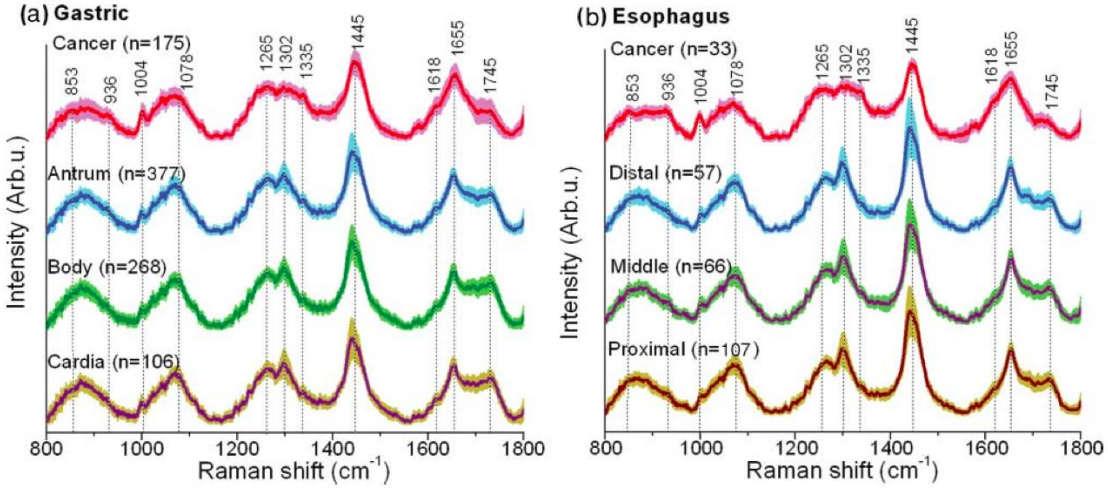


Figure 4.4 (a) *In vivo* mean Raman spectra ± 1 SD of normal (antrum, body, and cardia) and neoplastic gastric tissues. (b) *In vivo* mean Raman spectra ± 1 SD of the normal (distal, middle, and proximal) and neoplastic esophageal tissues. Note that the mean Raman spectra are vertically shifted for better visualization.

The Raman signals of the esophageal and gastric mucosa were reproducible among different patients. Prominent Raman bands are observed in both organs in the following peak positions with tentative biomolecular assignments (Table 1):[26-31, 89, 120] 853 cm^{-1} υ(C–C) of proteins, 936 cm^{-1} υ(C–C) of α-helix conformation for proteins, 1004 cm^{-1} $υ_s$(C–C) ring breathing of phenylalanine, 1078 cm^{-1} (υ(C–C) stretching mode of lipids), 1265 cm^{-1} Amide III υ(C–N) and δ(N–H) of proteins, 1302 and 1335 cm^{-1} δ(CH_2) deformation and CH_2CH_3 twisting of proteins and nucleic acids, 1445 cm^{-1} δ(CH_2) of proteins and lipids, 1618 cm^{-1} (υ(C=C) stretching mode of porphyrins), 1655 cm^{-1} Amide I υ(C=O) of proteins and 1745 cm^{-1} υ(C=O) of lipids.

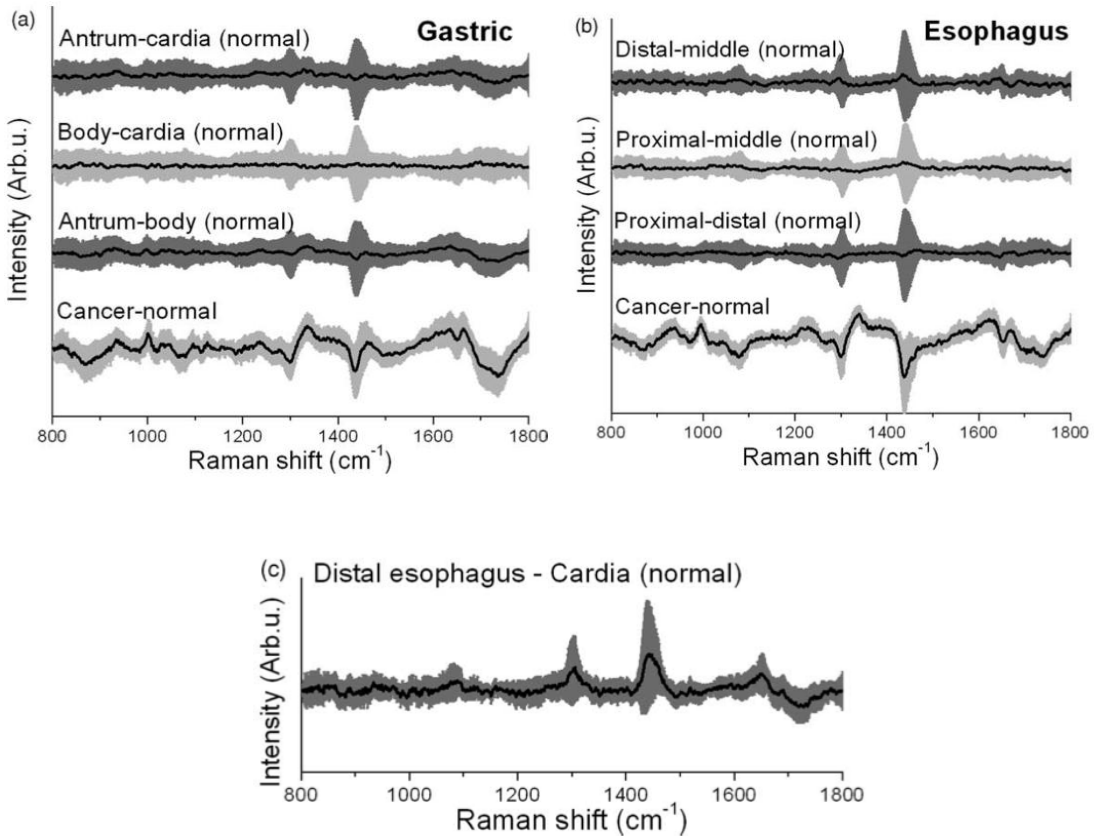


Figure 4.5 Difference spectra ± 1 SD comparing different anatomical regions and tissue types of the (a) gastric and (b) esophagus tissue. (c) Difference spectrum of the distal end of the esophagus and the cardia in the stomach, resolving inter-organ variability at the molecular level.

The difference spectra ± 1 SD between different tissue types (Figure 4.5a and Figure 4.5b) clearly resolve the subtle interanatomical variations and the major biochemical changes associated with cancer transformation; whereas Figure 4.5c reveals important biochemical differences between the two organs within the GE junction (i.e., the distal end of the esophagus and cardia in the stomach). One-way analysis of variance (ANOVA) with Bonferroni correction at 5%[28, 29] was separately employed for each organ (normal tissue). Table 4 summarizes the six significant Raman peak positions identified as well as the corresponding *p*-values generated from the post-hoc Fisher's least significant difference (LSD) test[28, 29] of all pair-wise anatomical regions in

normal gastric tissue (i.e., antrum versus body; body versus cardia; antrum versus cardia). Similar statistical analysis did not find significant interanatomical variability of normal tissues within the esophagus (i.e., distal, middle, and proximal); whereas significant spectral changes ($p < 1.0 \times 10^{-6}$) were found between the two organs (distal esophagus versus cardia), particularly at 1078, 1150, 1197, 1302, 1409, 1455, and 1745 cm^{-1}. One notes that no significant interanatomical variability ($p > 0.1$) of malignant tissues in the upper GI was found by Raman endoscope technique, which is probably due to the reason that the sizes of the tumor tissues measured in the esophagus and gastric are much larger than the probing volume (~1 mm^3) of our endoscopic Raman probe design.[64]

	p-values based on post-hoc Fisher's LSD test		
Raman peaks (cm^{-1})	Antrum versus body	Antrum versus cardia	Body versus cardia
936	$*9.1 \times 10^{-5}$	1.1×10^{-2}	4.4×10^{-1}
1245	$*2.0 \times 10^{-7}$	$*5.2 \times 10^{-5}$	8.9×10^{-1}
1335	$*1.1 \times 10^{-16}$	$*1.2 \times 10^{-10}$	2.1×10^{-2}
1618	$*6.7 \times 10^{-16}$	$*2.7 \times 10^{-3}$	9.5×10^{-2}
1665	$*6.7 \times 10^{-16}$	$*2.7 \times 10^{-3}$	9.5×10^{-2}
1745	$*3.1 \times 10^{-17}$	$*9.3 \times 10^{-12}$	4.0×10^{-1}

Table 4 The significance (*p*-value) of the distinctive Raman peaks identified (one-way ANOVA with Bonferroni correction at 5%) from *in vivo* normal gastric tissues (antrum, body, and cardia) based on the post-hoc Fisher's LSD test **p*-value < 0.01 (based on pair-wise comparison).

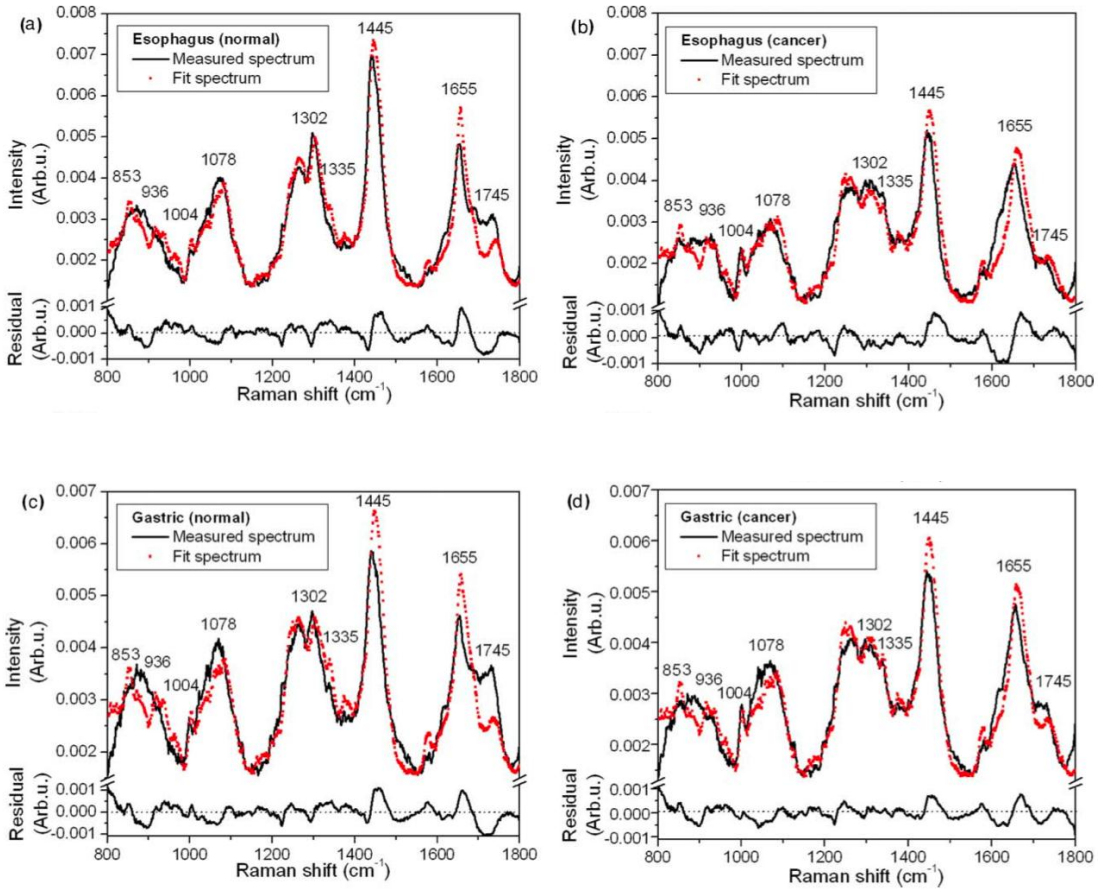


Figure 4.6 Comparison of the *in vivo* Raman spectra measured with the reconstructed tissue Raman spectra through the employment of the five basis reference Raman spectra: (a) normal esophagus, (b) esophageal cancer, (c) normal gastric, (d) gastric cancer. Residuals (measured spectrum minus fit spectrum) are also shown in each plot.

To characterize the spectral contribution from important Raman active tissue constituents (e.g., proteins, DNA, and lipids) in gastric and esophageal tissue types, the NNLSM was subsequently employed for semiquantitative modeling using the five basis reference spectra acquired as shown in Figure 4.3 (i.e., actin, collagen type I, DNA, histones, and triolein). Figure 4.6 shows the comparisons of the mean measured Raman spectra and the reconstructed Raman spectra of different tissues (i.e., gastric normal/cancer and esophageal normal/cancer). The residual variations of the NNLSM model rendered from each of the tissue sites in the gastric and esophagus are listed in Table 5. The fit-residuals between the reconstructed and the measured Raman spectra

are of less than 10%, indicating that the five tissue biochemical constituents used in NNLSM modeling can largely account for most information contained in tissue Raman spectra of the upper GI. To further assess the validity of biomolecular modeling for tissue characterization, we rendered difference spectra ± 1 SD of the reconstructed Raman spectra for estimating the similarities with the measured difference spectra. Figure 4.7a-b display the relative Raman spectral contribution of the five biochemicals of different tissue sites in the gastric and esophagus together with the corresponding *p*-value reflecting subtle interanatomical Raman spectral changes while significant differences associated with tissue carcinogenesis processes.

Tissue type	NNCLSM model residual variations
Gastric	
Antrum (normal)	$\pm 1.0 \times 10^{-3}$
Body (normal)	$\pm 1.0 \times 10^{-3}$
Cardia (normal)	$\pm 1.0 \times 10^{-3}$
Neoplasia	$\pm 1.0 \times 10^{-3}$
Esophagus	
Distal (normal)	$\pm 8.3 \times 10^{-4}$
Middle (normal)	$\pm 8.5 \times 10^{-4}$
Proximal (normal)	$\pm 8.8 \times 10^{-4}$
Neoplasia	$\pm 9.9 \times 10^{-4}$

Table 5 Residual variations in absolute numbers of the NNLSM model rendered from each of the tissue sites in the gastric and esophagus.

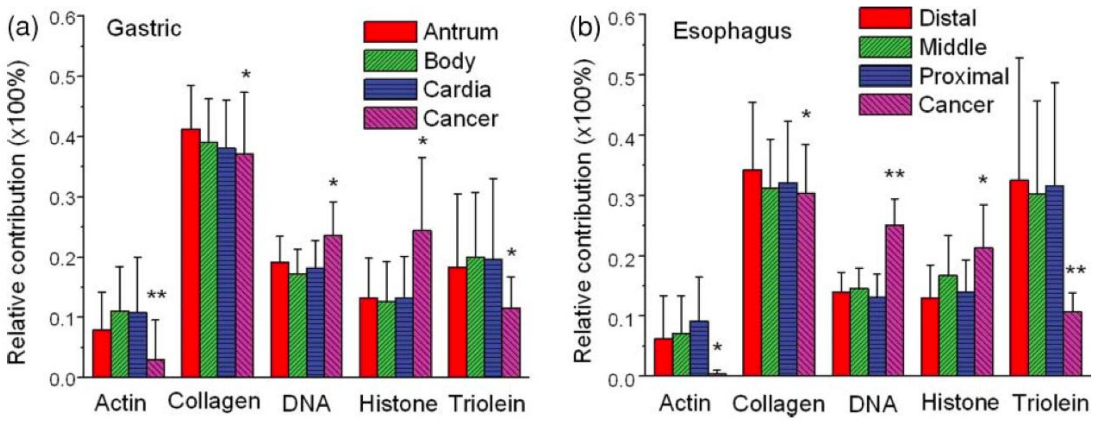


Figure 4.7 Histogram displaying the average biochemical spectral contributions to the different tissue types in the (a) gastric and (b) esophagus tissue. The one SD are shown for each model component. Note: * indicates a significant differences ($p < 1.0\times10^{-3}$); whereas ** indicates a significant differences ($p < 1.0 \times 10^{-8}$).

4.4 Discussion

In this chapter we developed a novel multimodal image-guided Raman endoscopic technique and reported on the *in vivo* gastric tissue characterization. With a specially designed miniaturized Raman endoscopic probe,[64] which can fit into the instrument channel of medical endoscopes for effective excitation and collection of tissue Raman scattered photons, the clinicians are now able to nondestructively assess the endogenous biochemical and morphological information of internal tissues under the wide-field endoscopic imaging (WLR, NBI, and AFI) guidance. Unlike *in vitro* Raman spectroscopy,[26-31] the image-guided Raman endoscopic technique can provide real-time biochemical assessment of tissues *in situ* (e.g., free of artifacts introduced by mechanically cutting, oxidation, hydration, loss of mucus layer, and vascular pressure, etc.) and therefore offer Raman endoscopy as a possible routine diagnostic tool in the esophagus and gastric. However, as the interanatomical regions of the esophagus and particularly the gastric tissue are highly biological and functionally specialized (e.g., thickness variability, distinct glandular types, secretion products, vascularity, etc.),[130,

[131] it is essential to figure out, to what extent, the effect of interorgan and interanatomical tissue sites' variations may affect the diagnostic algorithm developments of the Raman endoscope technique.

In this chapter, we evaluate for the first time *in vivo* Raman spectral properties of different anatomical regions of normal gastric and esophageal tissues. We have identified distinct Raman bands in both organs that are highly associated with proteins, DNA, and lipids (Figure 4.4). The difference spectra (Figure 4.5a-b) resolve the subtle interanatomical variability, indicating that the overall biomolecular and biochemical constituents of different normal tissues are very similar, whereas neoplastic tissues exhibit distinctive Raman spectral profiles (e.g., Raman peaks shifting, bandwidths broadening, or narrowing, relative Raman peak intensity changes). Further Raman spectral analysis also reveals significant spectral differences of tissues within the GE junction (Figure 4.5c), confirming the promising ability of Raman endoscopy for the sensitive characterization of morphological and biomolecular constituents in the upper GI tract. The Raman active biochemical profiles of gastric and esophageal tissue types were further assessed by rendering semiquantitative models constructed from essential biochemical constituents of gastric and esophageal tissues. For instance, proteins were found to be the most prominent contributors in *in vivo* tissue Raman spectra, which are similar to the *in vitro* tissue studies.[23, 94, 135] The difference spectra (Figure 4.5) also reflected large interpatient variations in the subregions 1078 υ(C–C) and 1440 cm^{-1} $\delta(CH_2)$ deformation, which are highly associated with signals from proteins and lipids, as confirmed by biochemical modeling. Moreover, the statistical analysis indicated subtle but complex differences in signals originating from proteins, lipids, and DNA of the antrum as compared to the body and cardia sites of the stomach. These

distinctive spectral features of the body likely reflect the intricate morphology[130] and biomolecular compositions of the body regions that constitute highly specialized cells (e.g., zymogenic and oxyntic cells in a densely packed architecture secreting acids, enzymes, etc.).[130] For instance, the relatively less collagen of the body site compared to antrum as revealed by biomolecular modeling ($p = 6.2\times10^{-2}$) could be associated with the closely packed morphologic structures containing less connective tissue. We also found that the *in vivo* Raman spectrum of cardia resembled the body region, which might be caused by the endoscopically poor defined and narrow zone of the cardia region and gradual transitional mucosa with cardia pathology-slides occasionally bearing resemblance with body histology (e.g., presence of oxyntic cell types).[130] But the difference spectra (Figure 4.5b) and statistical analysis indicated no statistically significant interanatomical variability of normal tissues within the esophagus ($p > 0.1$), confirming that the relatively thick squamous epithelium (~10–20 cell layers) is approximately identical among the distal, middle, and proximal regions of the esophagus.[131] Our Monte Carlo simulations[125] have indicated that the 785 nm laser light penetration depth from the Raman endoscopic probe developed is ~750–800 μm with a 200 μm beam size on the tissue surface, which is well within the lamina propria (the thickness of squamous epithelium is ~300–500 μm) of the esophagus.[131] Since most of the collected Raman scattered light originates from the shallower layer of the mucosa (e.g., epithelium) in which early neoplastic tissue transformation (e.g., dysplasia, carcinoma *in situ*) is most likely to originate, our fiber-optic Raman endoscopic probe still ensures a relatively small probing volume (of less than 1 mm^3 in tissue volume) covering the epithelium tissue layer, which permits the Raman evaluation of precancer and early cancer in the upper GI. On the other hand, the difference spectrum (Figure 4.5c) of normal distal esophagus and

gastric cardia within the GE junction exhibits significant Raman peaks shifting, bandwidths broadening/narrowing, and relative Raman peak intensity changes with the most prominent features at 1078, 1150, 1197, 1302, 1409, 1455, and 1715 cm^{-1}, while the NNLSM modeling suggests a complex variation in Raman spectral contributions of proteins DNA and lipids of different organs. These observations relate to the characteristic morphological architectures of the distinctive epithelial types (i.e., columnar lined versus nonkeratinized squamous epithelia) and glandular architectures etc. belonging to different organs[130, 131] as they were also visible under wide-field imaging observations (in particular NBI). The biochemical modeling indicates a considerably elevated content in lipids of the distal esophagus as compared to the gastric cardia ($p = 3.2\times10^{-6}$), which is in agreement with the reports on accumulations of lipids in the dense cell layers of the esophageal epithelium, as a part of the mucosal barrier.[136] Further, the relative less collagen of the esophagus compared to gastric is also linked with the thick squamous epithelium obscuring the Raman signal from the extensive connective tissues in the lamina propria. This optical filtering effect has also been observed in autofluorescence spectroscopy of gastric and esophageal neoplastic tissues as a result of the increased epithelial proliferation.[132, 137] It is particularly interesting that we observed an inverse linear relationship (correlation coefficient $R > 0.92$) between collagen and triolein in the esophagus, suggesting that the large interpatient variation in the lipids signals could be connected with epithelial properties (e.g., thickness, lipid content)[136] within the esophagus. To further evaluate the clinical implications of interorgan variability, we also employed a classification algorithm (e.g., PLS-DA)[90] to exploit subtle spectral variations of the entire Raman spectra that are not resolved by biomolecular modeling. The classification accuracy of 87.1% (142/163) could be achieved based on the leave-one

tissue site-out, cross-validation method,[90] proving that the Raman spectra reflect surface and subsurface tissue structures and biomolecular constituents. The differences in *in vivo* Raman spectra between neoplastic and nonneoplastic gastric tissues have also been reported in our previous gastric Raman studies.[89] The NNLSM modeling demonstrates that the neoplastic tissues show different contributions of distinctive biomolecules across interorgan and interanatomical regions in the upper GI, substantiating the important biochemical/biomolecular changes of tissues associated with carcinogenesis processes *in situ*. The significance of cancerous tissue transformation as compared to interanatomical variability in the gastric was further verified by ANOVA and post-hoc Fisher's LSD analysis (e.g., cancer versus normal gastric tissues (antrum, body, and cardia) with significant differences in the Raman bands: 936 ($p < 2.1\times10^{-2}$), 1245 ($p < 5.5\times10^{-4}$), 1335 ($p < 3.7\times10^{-28}$), 1618 ($p < 3.1\times10^{-31}$), 1665 ($p < 1.7\times10^{-25}$), and 1745 cm^{-1} ($p < 2.1\times10^{-30}$)). Particularly, the Raman peaks at 1575 and 1335 cm^{-1} are highly associated with increased DNA contents[122-124] that were also confirmed by biomolecular modeling (Figure 4.7); whereas the increased Raman band intensity and the bandwidth broadening of the 1655 cm^{-1} amide I υ(C = O) vibration pertained to a higher content of α-helical proteins (e.g., histones, the main protein component that makes up the chromatin). Overall, these biomolecular changes suggest the increased nuclear activity (e.g., hyperchromatic state) in neoplastic tissue, which is among the main characteristics of tissue carcinogenesis and is of considerable pathological value for tissue diagnosis and characterization. The biomolecular modeling also shows a small decrease in collagen, which is in agreement with the reports that cancerous cells proliferate, invade into underlying stromal layer and express a class of metalloproteases, resulting in an overall reduction of collagen content in cancer tissue.[123, 124] The relative increase

in Raman signals from proteins (e.g., 1004, 1265, and 1655 cm^{-1}) could also be related to increased mitotic activity,[137] while the reduction in Raman signals related to lipids (at 1078, 1302, 1440, and especially at 1745 cm^{-1}) for precancer and cancer gastric tissues has also been found in our previous Raman studies.[78, 122, 132] On top of these, the Raman peak increase at 1575–1618 cm^{-1} is highly associated with angiogenesis process of neoplastic tissue,[78, 122, 132] which in fact, is reflected by the poor fitting ability of the NNLSM model in this spectral region due to the omission of Raman-active blood constituents (e.g., hemoglobin) in the basis reference set. Figure 4.7a-b uncover the diagnostic significant fit coefficients for detection of neoplasms in the gastric and esophagus (e.g., actin, collagen, DNA, histones, and triolein); and the consistency in identifying similar diagnostic significant biochemicals across interorgans and interanatomical regions reconfirms the versatility of the Raman endoscopy for early detection and diagnosis of cancers based on Raman signals derived from proteins, DNA, and lipids. To further evaluate the clinical implications of interorgan variability, we also employed PLS-DA for cancer diagnosis whereby a diagnostic accuracy of 89.3% (sensitivity: 92.6% (162/175); specificity 88.6% (665/751)) and of 94.7% (sensitivity: 90.9% (30/33); specificity 93.9% (216/230)) in the gastric and esophagus, respectively, can be achieved based on the leave-one tissue site-out, cross validation method. The Raman spectral changes associated with cancerous tissue transformation (Figure 4.5) were several orders of magnitude larger than interanatomical variability of normal tissue, illustrating the efficacy of *in vivo* Raman endoscopy technique for early cancer diagnosis and detection in the upper GI.

4.5 Conclusion

High-quality *in vivo* Raman spectra can be acquired from different sites of the upper GI (i.e., gastric and esophagus) in real-time by Raman endoscopy technique

during clinical endoscopic examinations. Difference spectra together with biochemical modeling suggest that interorgan variability is significant whereas the interanatomical variability of the esophagus and gastric is subtle compared to neoplastic tissue transformation.

Chapter 5. Fiber-optic Raman spectroscopy probes gastric carcinogenesis *in vivo* at endoscopy

Intestinal-type gastric carcinogenesis is a complex multi-step disease, and early precursors (e.g. IM and dysplasia) can be very challenging to identify using conventional white-light endoscopic imaging. In this chapter we assess the capability of the developed Raman spectroscopy technique for multi-class elucidation of intestinal-type gastric carcinogenesis sequence *in vivo* at endoscopy. We analyze 1277 *in vivo* Raman spectra from 83 gastric patients associated with intestinal-type carcinogenesis. The results show that *in vivo* Raman spectroscopy integrated with semi-quantitative spectral modeling (e.g. DNA, lipids, glycoprotein, proteins and blood) reveals the progressive changes of biochemical constituents in gastric tissue associated with preneoplastic and neoplastic transformation (i.e., IM, dysplasia and adenocarcinoma). Multi-class probabilistic partial least squares-discriminant analysis (PLS-DA) diagnostic algorithms based on *in vivo* Raman spectra are able to identify normal mucosa with sensitivity of 75.88% and specificity of 87.21%; IM with sensitivity of 46.67% and specificity of 87.55%; dysplasia with sensitivity of 83.33%; specificity of 95.80%, and adenocarcinoma with sensitivity of 84.91% and specificity 95.57%, respectively. This work demonstrates that Raman spectroscopy is a sensitive biomolecular probe for monitoring intestinal-type gastric carcinogenesis *in vivo* and realize diagnosis of precancer and cancer during clinical endoscopic examination.

5.1 Introduction

Intestinal-type gastric cancer is a result of a multi-step disease pathway that is known to be associated with the intermediate transformation into IM phenotype before the progression to dysplasia and invasive carcinoma.[1, 3, 138] Early identification and diagnosis of precancer and other precursors (e.g. IM) together with adequate treatment (e.g. endoscopic submucosal dissection (ESD) or HP eradication) is a critical measure to improving the survival rates of the gastric patients.[3] However, endoscopic identification of IM, dysplasia and early flat cancer remain very difficult using conventional WLR imaging or NBI that heavily relies on gross morphological and microvascular manifestations at clinical endoscopy.[19, 139] Our preceding NIR Raman spectroscopy studies of stomach tissue *ex vivo* have established the diagnostic foundation for detection and diagnosis of different types of pathologic gastric tissues (e.g. HP infection, IM, dysplasia, neoplasia, diffuse and intestinal-type adenocarcinoma) with good accuracies (of ~85 – 96%).[26, 28-30] The clinical potential of Raman spectroscopy for multi-class identification of different stages in the gastric carcinogenesis sequence (for example IM versus dysplasia or dysplasia versus neoplasia), however, has not been evaluated in detail to date. The discrimination between different stages/types of gastric carcinogenesis is of particular clinical relevance, which may influence the decision of treatment strategies (e.g. ESD or surgery) for gastric patients. For instance, IM typically only requires HP eradication whereas endoscopic resection is more applicable to dysplastic lesions. Cancer on the other hand may require partial or full gastrectomy. Hence, the main aim of this chapter is to examine the capability of Raman endoscopy for the direct assessment of intestinal-type gastric carcinogenesis sequence *in vivo* during clinical endoscopic examination.

5.2 Materials and methods

5.2.1 Raman instrumentation

The Raman spectroscopy system used for *in vivo* tissue Raman measurements has been reported in detail in Section 4.2.1.

5.2.2 Data processing

The data processing has previously been reported in detail in Section 4.2.2.

5.2.3 Study protocol

The patient recruitment and ethical protocol has been reported in detail in Chapter 3. In this work, Raman data from 83 patients were analyzed including spectra from normal tissue (n=850), IM (n=71), dysplasia (n=92) and intestinal type adenocarcinoma (n=264). Tissue histopathology serves as the gold standard for evaluation of the performance of Raman technique for *in vivo* tissue diagnosis and characterization.

5.2.4 Spectral modeling

To assess the semi-quantitative differences between different tissue types *in vivo*, we have developed an on-line spectral modelling of the *in vivo* tissue Raman spectra using basis reference biochemical.[132] The semi-quantitative spectral modeling technique has been described in Section 2.4.5 Statistical significance (*p*-values) was calculated using one-way ANOVA with Fisher post hoc least significant difference (LSD) test at the 0.05 level.

5.2.5 Multi-class diagnostics

To realize *in vivo* discrimination between different gastric tissue pathologies, the multi-class probabilistic PLS-DA is utilized as described in Section 2.4.3. Leave-one

patient-out cross-validation was used to assess PLS-DA model complexity. We further extend the dichotomous PLS-DA algorithm for multi-class classification using one-against-one scheme with max-wins voting.[82] Thus, a total of 6 PLS-DA models were rendered (e.g. normal vs. IM; normal vs. dysplasia etc.) with varying complexities spanning from 1 to 3 LVs. Independent validation in this study was then performed by splitting the data-set into training (~80% of dataset) and testing (~20% of dataset). The split testing was performed on the patient basis ensuring independent validation.

5.3 Results

A total of 1277 *in vivo* tissue Raman spectra (normal gastric mucosa (n = 850), IM (n=71), dysplasia (n=92), intestinal-type adenocarcinoma (n=264)) were analyzed from 237 tissue sites of 83 gastric patients for gastric tissue diagnosis and characterization. Figure 5.1 shows the mean Raman spectra ±1 SD of normal gastric mucosa, IM, dysplasia and intestinal-type adenocarcinoma.

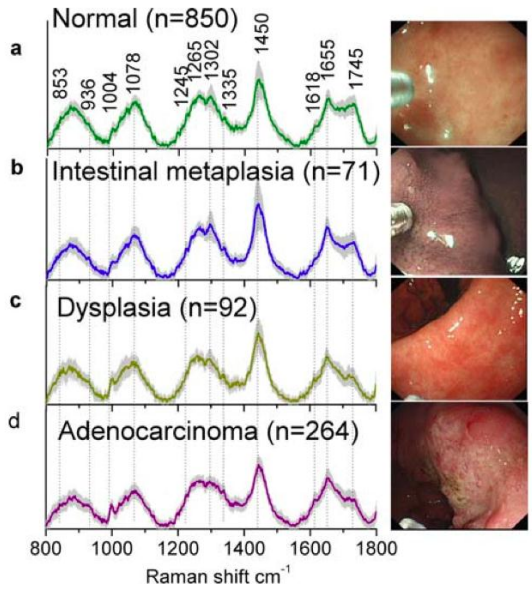


Figure 5.1 Mean *in vivo* Raman spectra ±1 standard deviation (SD) of (a) normal mucosa; (b) intestinal metaplasia; (c) dysplasia, and (d) intestinal-type adenocarcinoma. Also shown are the corresponding Raman endoscopy procedures under white light reflectance (WLR) and narrow-band imaging (NBI) guidance.

The corresponding images of the associated WLR/NBI guided Raman endoscopy procedures are also shown in Figure 5.1. The spectral differences (e.g. peak intensity and bandwidth) between different tissue types are subtle, but are consistent over most parts of the Raman spectra ($p < 0.05$). The prominent tissue Raman peaks (Table 1) are identified with tentative biomolecular assignments[27, 64, 97, 122] as follows: 936 cm^{-1} (υ(C-C) proteins), 1004 cm^{-1} (υ(C–C) ring breathing of phenylalanine), 1078 cm^{-1} (υ(C–C) of lipids), 1265 cm^{-1} (amide III υ(C–N) and δ(N–H) of proteins), 1302 cm^{-1} (CH_3CH_2 twisting and wagging of proteins and lipids), 1445 cm^{-1} (δ(CH_2) deformation of proteins and lipids), 1655 cm^{-1} (amide I υ(C=O) of proteins) and 1745 cm^{-1} (υ(C=O) of lipids). Figure 5.2 gives the representative hematoxylin and eosin (H & E) slides (×200 magnification) of the corresponding tissue sites in intestinal-type gastric carcinogenesis measured using Raman endoscopy, which include (a) normal oxyntic gastric mucosa; (b) extensive IM where the gastric epithelium contains apparent goblet cells; (c) low-grade dysplasia showing abnormal nuclear content and the associated intestinal metaplasia, and (d) intestinal-type gastric adenocarcinoma.

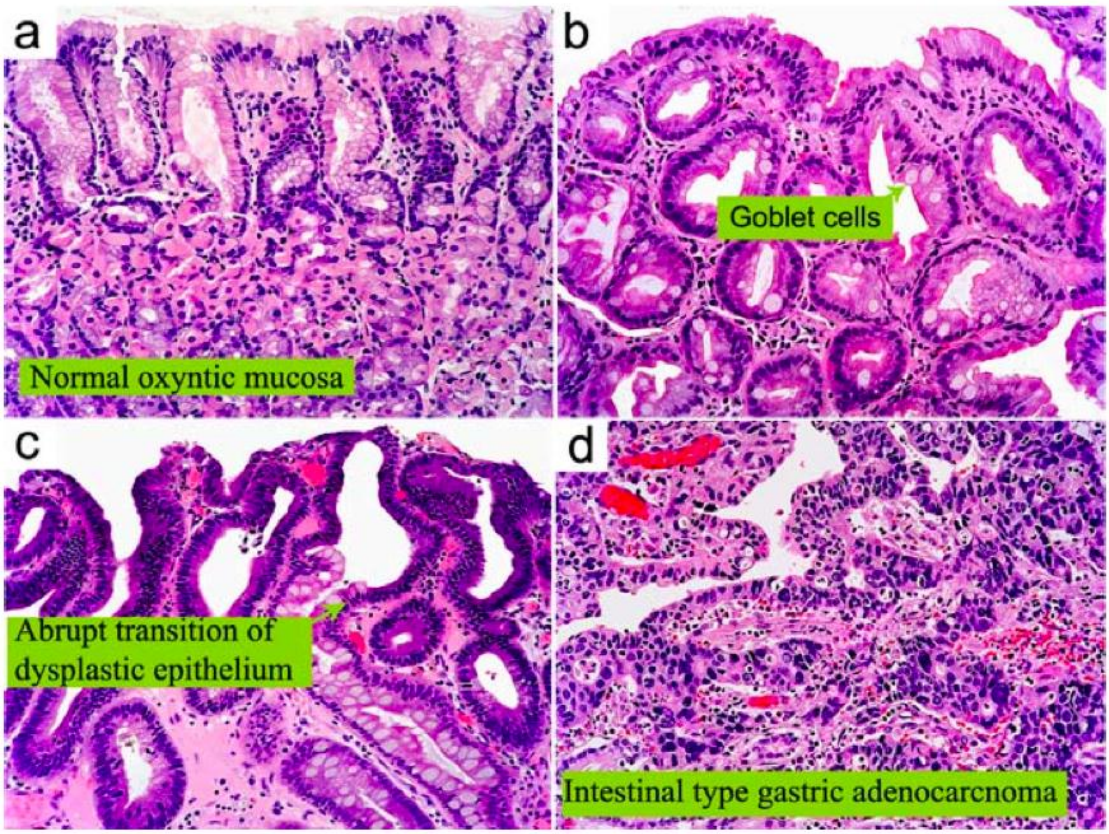


Figure 5.2 Photomicrographs of the haematoxylin and eosin (H&E)-stained sectioned slides (×200 magnifications) of gastric tissues: (a) normal oxyntic mucosa; (b) extensive intestinal metaplasia; (c) low-grade dysplasia with associated intestinal metaplasia; (d) intestinal-type adenocarcinoma.

Semi-quantitative spectral modelling was used to correlate spectral changes with the biomolecular and histopathological manifestations in gastric intestinal- type carcinogenesis (Figure 5.2). Figure 5.3a shows the measured Raman spectra of the dominant reference basis components (i.e. proteins, glycoprotein, DNA, lipids and blood) in gastric tissue. The measured tissue Raman spectra (black color) and reconstructed spectra (red color) (Figure 5.3b) agree well with each other (fit residuals of <10%), suggesting the important role of these biochemical components in gastric tissue. Figure 5.4a-e summarizes the relative spectral contributions ±1 SE in bar diagrams together with the corresponding *p*-values. The results show a gradual transformation from normal to IM, dysplasia and ultimately adenocarcinoma, reflecting the progressive accumulation of biomolecular and morphological changes

in the gastric tissue. Of particular significance, dysplastic and cancerous lesions showed progressive changes in DNA ($p < 0.01$), blood ($p < 0.001$) and lipid content. The above results demonstrate that Raman endoscopy can directly monitor the multi-step gastric carcinogenesis sequence *in vivo* ranging from normal tissue, intestinal metaplasia, dysplasia and final progression into adenocarcinoma exclusively based on biomolecular manifestations. To exemplify it graphically, Figure 5.5a depicts a scatter plot of blood versus DNA content for normal + IM, dysplasia and adenocarcinoma, illustrating the diagnostic significance of nuclear abnormalities and neovasculation. Figure 5.5b shows the proteins versus lipids content for dysplastic and cancerous lesions, elucidating the distinct clustering between precancer and cancer tissue. Based on these specific endogenous biomolecules, neoplasia can be discriminated from dysplasia with a sensitivity of 76.73% and specificity of 61.96%. Overall, the above results substantiate that Raman endoscopy probes the intestinal-type gastric carcinogenesis sequence and can aid in objective histopathological assessments of gastric tissue *in vivo* based on the intrinsic biomolecular signatures.

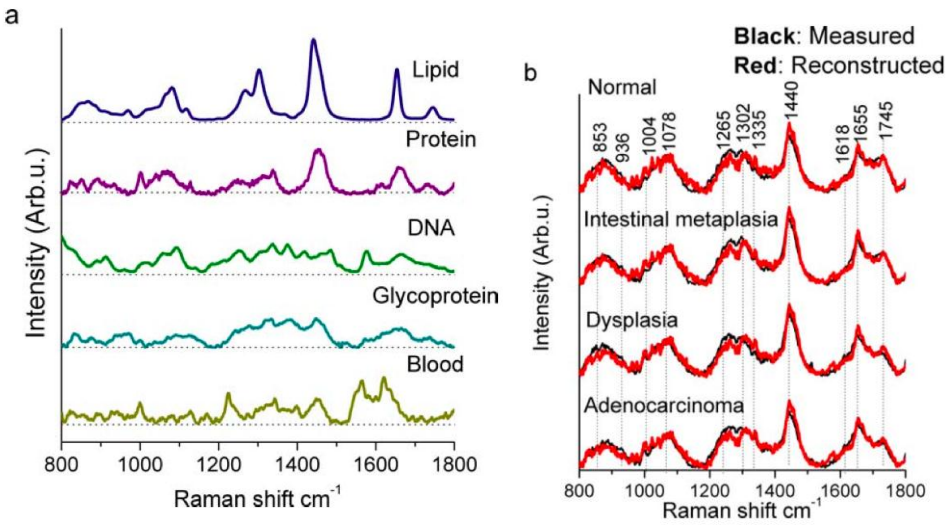
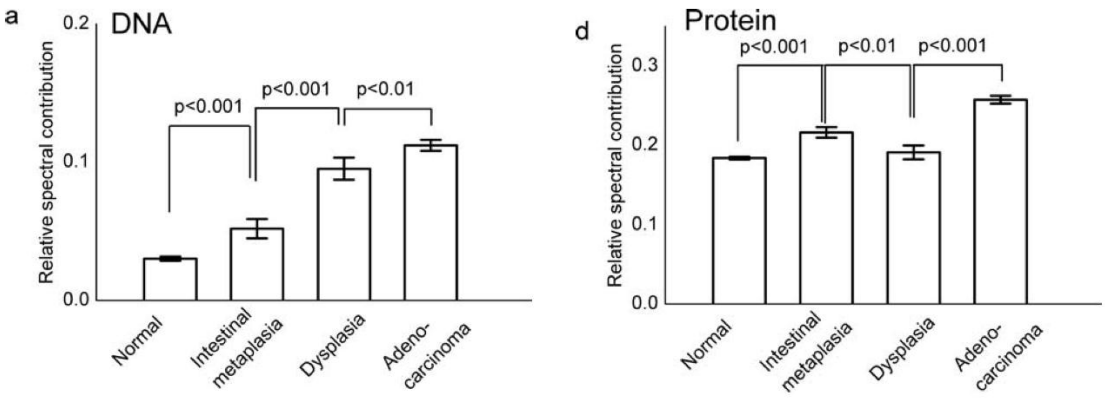


Figure 5.3 (a) The five basis reference Raman spectra from lipids, DNA, glycoprotein, proteins and blood used for Raman biomolecular modelling of carcinogenesis process. (b) Comparison of the measured *in vivo* gastric Raman spectra (black color) with the reconstructed tissue Raman spectra (red color) through Raman biomolecular modelling among different gastric tissue types (i.e. normal mucosa, intestinal metaplasia, dysplasia and adenocarcinoma).

To realize multi-class discrimination between different histopathological types, we rendered probabilistic PLS-DA algorithms for differentiation among the four gastric tissue types (i.e. normal, IM, dysplasia and adenocarcinoma) by splitting the dataset into training (~80% of dataset) and testing (~20% of dataset). Table 6 summarizes the probabilistic PLS-DA prediction results (i.e. sensitivity, specificity and accuracy) of the testing dataset in a confusion matrix. We found that normal mucosa could be detected with accuracy of 79.69%, (sensitivity: 75.88%; specificity: 87.21%). IM phenotype was identified with accuracy of 85.16% (sensitivity: 46.67%; specificity: 87.55%), whereas dysplastic tissues were identified with accuracy of 94.92% (sensitivity: 83.33%; specificity: 95.80%) and adenocarcinoma with accuracy of 93.36% (sensitivity: 84.91%; specificity: 95.57%). These findings suggest that besides resolving gastric intestinal-type carcinogenesis at the biomolecular level (Figure 5.4), Raman endoscopy can also be used for targeted biopsies of high-risk lesions as well as discrimination between precancer and cancer during clinical endoscopic inspections.



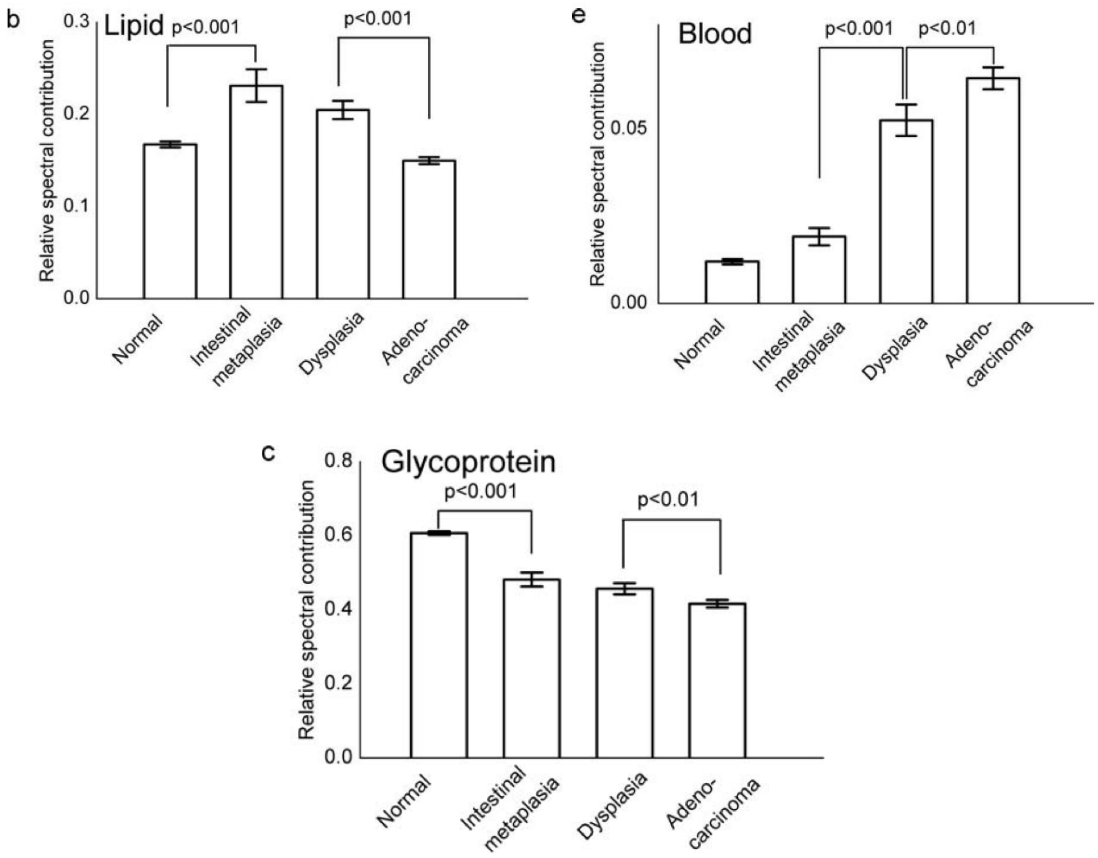


Figure 5.4 Bar histograms ~1 standard error (SE) of the relative fit coefficients of (a) DNA; (b) mucin; (c) lipids; (d) proteins and (e) blood for the four different tissue types (i.e. normal mucosa, intestinal metaplasia, dysplasia and intestinal- type adenocarcinoma).

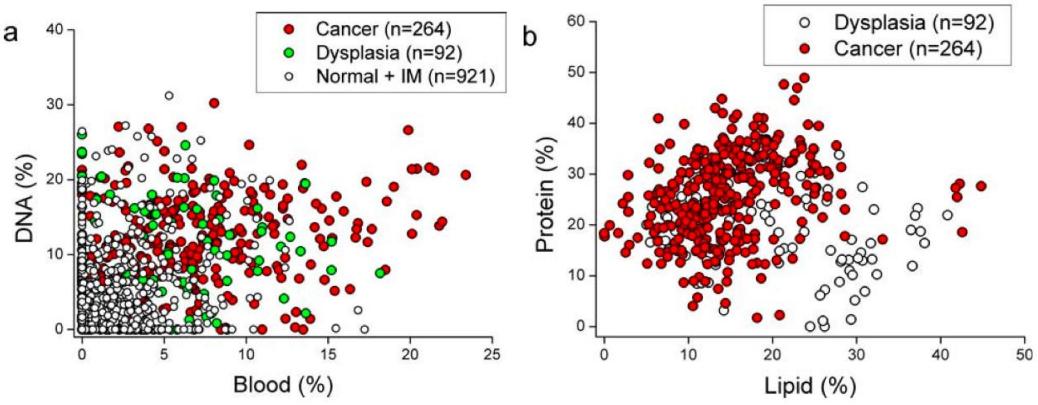


Figure 5.5 (a) Scatter plots of blood versus DNA contents for normal + IM, dysplasia and intestinal-type adenocarcinoma. (b) Lipids versus proteins contents for dysplasia and intestinal-type adenocarcinoma.

Raman prediction				
Histopathology	Normal	Intestinal metaplasia	Dysplasia	Adenocarcinoma
Normal	129	30	4	7
Intestinal metaplasia	7	7	0	1
Dysplasia	2	0	15	1
Adenocarcinoma	2	0	6	45
Sensitivity (%)	75.88	46.67	83.33	84.91
Specificity (%)	87.21	87.55	95.80	95.57
Accuracy (%)	79.69	85.16	94.92	93.36

Table 6 Confusion matrix and diagnostics of ~20% of the dataset (n=256 spectra) based on *in vivo* Raman spectroscopy and multi-class probabilistic partial least squares-discriminant analysis (PLS-DA) together with one-against-one voting.

5.4 Discussion

The pathogenesis of intestinal-type gastric adenocarcinoma is a complex cascade process of biomolecular and morphological changes that include intermediate transformation to IM before development of dysplasia and the risk leading to invasive carcinoma (Figure 5.2).[138] Identification of early predisposing lesions is very challenging during clinical endoscopic procedures. Thus, the development of a label-free optical diagnostic technology that can interrogate the biomolecular information of tissue architecture and cytology *in vivo* would be of great clinical value at endoscopy. The distinction between normal, precancer and cancer in intestinal-type gastric carcinogenesis in this study is of highly clinical relevance, and could greatly reduce multiple redundant biopsies to tremendously benefit the gastric patients at high risk.

This work is the first fiber-optic Raman endoscopic study to investigate the multi-step intestinal-type carcinogenesis sequence *in vivo* in the stomach time during clinical gastroscopy. The spectral modelling based on endogenous biomolecular components was able to derive semi-quantitative biomolecular information from the *in vivo* tissue Raman spectra with subtle fit residuals (of <10%), despite of the gross simplification without modelling of inter- and intra-molecular interactions, optical absorption and scattering properties of tissue, etc.[122] Remarkably, Raman endoscopy

together with spectral modelling resolves the cascade of biomolecular events, and tissue Raman features identified correlate well with the histopathology of gastric carcinogenesis (Figure 5.2). In particular, we found that IM showed an increased content of DNA ($p < 0.001$) and proteins ($p < 0.001$) compared to normal mucosa. This observation is consistent with flow cytometric studies revealing DNA aneuploidy in IM that is characterized by an increased hyperproliferative activity of gastric epithelial cells.[140] With the assumption that different mucin subtypes produce very subtle spectral differences,[23] it is plausible that the overall mucin content can be reflected in the spectral modeling. The changes in glycoprotein content is likely caused by the transition to a glandular phenotype expressing goblet, paneths and absorptive cells for IM pathology that secrete mucins (i.e. sialomucins and sulfomucins)[27] different from those secreted by the normal gastric foveolar cells (e.g. neutral mucin such as MUC5AC). The changes in glycoprotein associated with carcinogenesis ($p < 0.05$) could be explained with variations and losses in expression of distinct mucin subtypes (e.g. MUC2, MUC5AC, and MUC6). Neoplastic tissues were also highly associated with up-regulated DNA ($p < 0.001$) and proteins ($p < 0.01$) contents in accordance with histopathological and cellular characterization (Figure 5.2). Overall, the semi-quantitative spectral modelling indicates a gradual disease transformation from normal, IM to dysplasia and ultimately adenocarcinoma that likely reflect a multitude of oncogenes, cell cycle regulators, growth factors (i.e. progressive accumulation of biochemical and biomolecular changes)[116, 137, 140-142] together with the changes in tissue architecture. In addition, we also observe a prominent reduction in lipid Raman signals in neoplastic tissue ($p < 0.001$) that is likely due to the thickening of mucosa and light absorption obscuring the Raman signals from deeper subcutaneous fat layers. Further, adenocarcinoma tissue shows

intense Raman peaks at 1004 cm^{-1} (phenylalanine υ_s(C–C)), amide III (1265 cm^{-1}) and amide I band (1655 cm^{-1}) that are optical biomarkers related to increased protein contents. These results suggest the increased proliferation of cancer cells constituting the gastric tissue.[137] The spectral modelling (Figure 5.4) also confirms that Raman spectroscopy resolves the gradual angiogenesis and resulting neovasculation processes of gastric carcinogenesis *in vivo*. Overall, *in vivo* Raman endoscopy in conjunction with semi-quantitative spectral modelling provides good agreement with histopathological characterization in gastric tissue *in vivo* at the molecular level. Our previous Raman studies[26-29] showed that Raman endoscopy yielded a good diagnostic accuracy (>90%) for *in vivo* diagnosis of gastric dysplasia and neoplasia. In this work, our probabilistic PLS-DA multi-class analysis discloses that Raman endoscopy can be further used for targeted biopsies and discrimination between dysplastic and neoplastic lesions (Figure 5.5a). The finding reconfirms the capability of fiber-optic Raman spectroscopy to directly assess the depth of invasion of lesions occurring either in the epithelium tissue or bulky malignant lesions. Misclassifications in Table 6 are likely caused by the diffuse histopathological borderline between high-grade dysplasia and cancer that might be partially solved with consensus pathology. On the other hand, the sensitivity of detecting IM was rather poor, which may be due to the reason that this volume-type Raman probe interrogates with gastric tissue of up to 800 μm in depth, diluting the subsurface (IM) tissue Raman signal with deeper bulky tissue Raman. The current volume-type Raman probe design may not be efficient for selectively targeting IM that is largely confined to the epithelium. With the development of the second generation of confocal Raman probe designs for depth-resolved measurements currently under testing in the Optical Bioimaging Laboratory at NUS, the confocal Raman spectroscopy is expected to increase the detection

sensitivity for both IM and early dysplasia by selectively measuring tissue signals from the epithelium. Nevertheless, the biomolecular spectral modeling on *in vivo* Raman spectra in this work gives rise to statistical significant differences in the Raman spectra among IM phenotype and normal mucosa. *In vivo* Raman endoscopy reveals the specific endogenous biomolecules (i.e. optical biomarkers) responsible for the progression of intestinal-type gastric carcinogenesis *in situ*. Raman endoscopic technique can discriminate between normal mucosa, dysplasia and invasive cancer with high biomolecular specificity. Diagnostic accuracies of 80 – 95% are adequate to guide the targeted biopsy of suspicious lesions, thereby reducing the number of negative biopsies. For instance, dysplastic cell can eventually appear within areas of IM. As dysplastic foci are prone to sampling errors during endoscopy, Raman spectroscopy can benefit by accurately identifying these high-risk lesions for targeted biopsies. One notes that the current *in vivo* gastric tissue Raman datasets acquired are unbalanced with a relatively larger number of true negatives but a smaller pathologic tissue (e.g. IM, dysplasia), which may give rise to a biased overall accuracy within the PLS-DA models. Currently, clinical Raman measurements on a larger series of gastric patients to obtain balanced numbers of different pathologic groups are in progress in the Endoscope Centre at NUH for further evaluating the true clinical merits of Raman endoscopic technique for improving the early diagnosis of gastric precancer and cancer during endoscopic examination.

5.5 Conclusion

This chapter assessed the capability of Raman spectroscopy for multi-class elucidation of intestinal-type gastric carcinogenesis sequence *in vivo* for improving precancer detection at endoscopy. We employ the novel image-guided Raman endoscopy technique developed for *in vivo* gastric tissue Raman measurement within

0.5 s during clinical endoscopic examination. *In vivo* Raman spectroscopy integrated with semi-quantitative spectral modeling (e.g. DNA, lipids, glycoprotein, proteins and blood) reveals the progressive changes of biochemical constituents in gastric tissue associated with preneoplastic and neoplastic transformation (i.e., IM, dysplasia and adenocarcinoma). In this chapter we demonstrate that Raman spectroscopy is a sensitive biomolecular probe for monitoring intestinal-type gastric carcinogenesis to realize early diagnosis and detection of precancer and early gastric cancer *in vivo* during clinical endoscopic examination.

Chapter 6. Development of a real-time framework for *in vivo*, on-line diagnosis during clinical endoscopic examination

Data analysis of tissue Raman spectra in the literature has mostly been limited to post-processing and off-line algorithm development. In this Chapter, a comprehensive automated on-line Raman spectral diagnostics framework is developed to realize real-time prospective *in vivo* cancer detection at endoscopy. A total of 2748 *in vivo* gastric tissue spectra (2465 normal and 283 cancer) were used from 305 patients recruited to construct a spectral database for diagnostic algorithms development. The novel diagnostic scheme developed implements on-line preprocessing, outlier detection based on principal component analysis statistics (i.e., Hotelling's T^2 and Q-residuals) for tissue Raman spectra verification as well as probabilistic multivariate diagnostics. Free-running optical diagnosis with multivariate/chemometric processing times of < 0.13 s can be achieved, which is important to realizing real-time *in vivo* tissue diagnostics during clinical endoscopic examination. The optimized partial least squares-discriminant analysis (PLS-DA) models based on the randomly resampled training database (80% for learning and 20% for testing) provide the diagnostic accuracy of 85.6% (95% confidence interval (CI): 82.9% to 88.2%) (sensitivity of 80.5% (95% CI: 71.4% to 89.6%) and specificity of 86.2% (95% CI: 83.6% to 88.7%)) for the detection of gastric cancer. The PLS-DA algorithms are further applied prospectively for the first time on 10 gastric patients at gastroscopy, achieving the predictive accuracy of 80.0% (60⁄75) (sensitivity of 90.0% (27⁄30) and specificity of 73.3% (33⁄45)) for *in vivo* diagnosis of gastric cancer. This chapter successfully moves biomedical Raman spectroscopic technique into real-time, on-line prospective clinical cancer diagnosis.

6.1 Introduction

Few groups have reported on the implementation of real-time tissue Raman spectroscopy.[143] Feld et al. presented a synchronized laser Raman system that could acquire and process Raman spectra within 2 s using spectral modeling of reference biochemicals with an aid of least squares regression techniques.[93, 144-148] Zeng et al. developed a real-time integrated Raman system for evaluating skin that could acquire and process the signal within 1.1 s.[143] Nevertheless, most of the current Raman spectroscopic analyses have been limited to off-line post processing for classification of spectra with cross-validation procedures, which render practical limitations including the setting of exposure times, post-verification of spectrum quality, and lack of automatic feedback mechanisms to the clinicians for implementation of straight-forward probabilistic diagnostics in clinical settings. Fully automated tissue spectral quality verification and real-time tissue cancer diagnostics are vital to translating the Raman spectroscopic diagnostic technique into prospective clinical endoscopic routine. In this work, we develop an on-line biomedical spectral diagnostic framework integrated with the novel image-guided Raman endoscopy for real-time probabilistic detection of cancer in the upper GI. We also validate the efficacy of the on-line Raman framework developed for prospective real-time prediction in patients with gastric malignancies at clinical endoscopy.

6.2 Materials and methods

6.2.1 Raman instrumentation

The Raman spectroscopy system used for *in vivo* tissue Raman measurements has been reported in detail in Chapter 4.2.1.

6.2.2 Study protocol

The ethical protocol has been reported in Chapter 3. In this work, a total of 2748 *in vivo* gastric tissue spectra (normal (n=2465) and cancer (n=283)) from the 305 patients were used to construct the spectral database for developing diagnostic algorithms for gastric cancer diagnostics. Tissue histopathology serves as the gold standard for evaluation of the performance of Raman technique for *in vivo* tissue diagnosis and characterization.

6.2.3 Development of on-line Raman spectroscopic framework

The on-line biomedical Raman spectroscopic framework developed has been implemented as a graphical user interface (GUI) under the Matlab 2011a (Mathworks Inc., Natick, MA) scripting environment in a fast computing workstation (64 bit I7 quad-core 4GB memory). This framework has been thoroughly optimized for rapid data processing for real-time tissue diagnostics. Hardware components of the rapid Raman system (e.g., laser power control, spectrometer, CCD shutter and camera readout synchronization) have been interfaced to the Matlab software through libraries for different spectrometers/cameras (e.g., PVCAM library (Princeton Instruments, Roper Scientific, Inc., Trenton, NJ) and Omni Driver (Ocean Optics Inc., Dune- din, FL), etc.). A schematic of the spectral acquisition and processing flow of on-line diagnostic framework is depicted in Figure 6.1. The laser was electronically synchronized with the CCD shutter. The automatic adjustment of laser power, exposure time and accumulation of spectra were realized by scaling to within 85% of the total photon counts (e.g., 55,250 of 65,000 photons) based on preceding tissue Raman measurements, whereas an upper limit of 0.5 s was set to realize clinically

acceptable conditions. The accumulation of multiple spectra and automatic adjustment of exposure time provides a rapid and straightforward methodology to prevent CCD saturation and to obtain high signal-to-noise ratios (SNR) for endoscopic applications. The Raman-shift axis (wavelength) was calibrated using a mercury/argon calibration lamp (Ocean Optics Inc., Dunedin, FL). The spectral response correction for the wavelength-dependence of the system was conducted using a standard lamp (RS-10, EG&G Gamma Scientific, San Diego, CA). The reproducibility of the platform can be continuously monitored with the laser frequency and Raman spectra of cyclohexane and acetaminophen as wavenumber standards. All the system performance measures including CCD temperature, integration time, laser power, CCD alignment are accordingly logged into a central database via SQL server. Due to the inter-anatomical and inter-organ spectral variances as we observed (Chapter 4), the on-line framework we designed implements organ-specific diagnostic models in a database structure and can instantly switch among the spectral databases of different organs (e.g., esophagus, gastric, colon, cervix, bladder, lung, nasopharynx, larynx, and the oral cavity (hard palate, soft palate, buccal, inner lip, ventral and the tongue)), making this Raman platform a unified diagnostic tool for cancer detection.

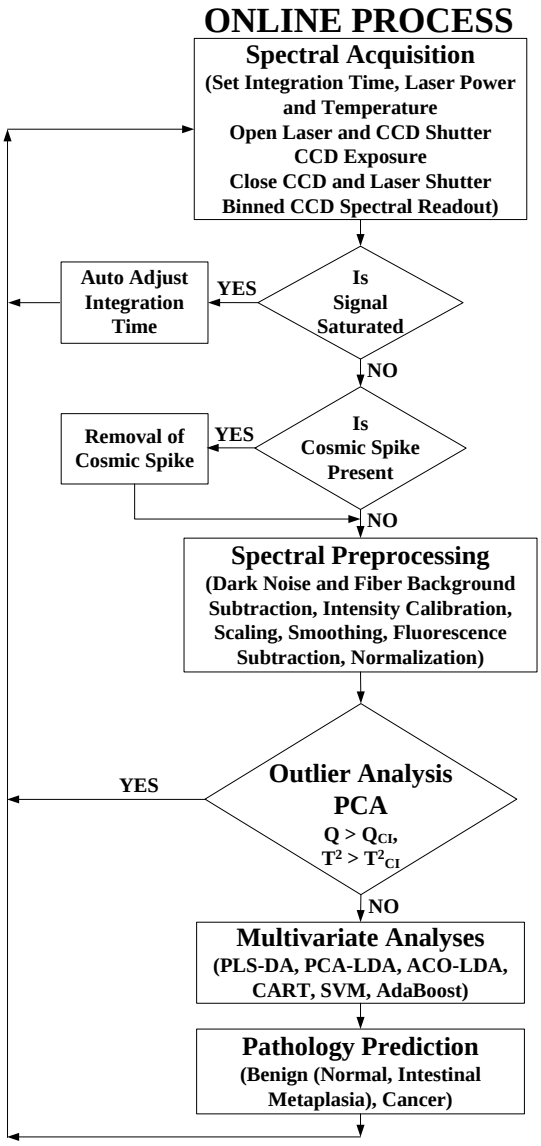


Figure 6.1 Flow chart of the on-line biomedical Raman spectroscopic diagnostic acquisition framework.

6.2.1 On-line preprocessing and outlier detection

Real-time preprocessing of Raman signals was realized with the detection of cosmic rays using the first derivative with a 95% or 99% CI over the whole spectral range set as a maximum threshold. Data points lying outside of a threshold were interpolated to 2nd order. The spectra were further scaled with integration time and

laser power. A Savitzky-Golay smoothing filter was used to remove noise in the intensity corrected spectra, while a fifth order modified polynomial constrained to the lower bound of smoothed spectra was subtracted to resolve the tissue Raman spectrum alone.[25, 77, 117] The Raman spectra were further normalized to the integrated area under the curve, enabling a better comparison of the spectral shapes and relative Raman band intensities among different tissue pathologies. The spectra were then locally mean-centered according to the specific database to remove common variations in the data. Following preprocessing, the Raman spectra were fed to a model-specific outlier analysis.

We incorporate an on-line outlier detection scheme into biomedical spectroscopy as a high-level model-specific feedback tool in our on-line framework by using PCA coupled with Hotelling's T^2 and Q-residual statistics.[149, 150] Briefly, PCA reduces the dimension of the Raman spectra by decomposing them into linear combinations of orthogonal PCs, such that the spectral variations in the dataset are maximized. The PCA model is defined by Equation [4]:[151] Accordingly, Hotelling's T^2 statistics is a measure of variance captured by the PCA model (sample to model distance) and is defined by:[149]

$$T_{ik}^2 = t_{ik}(\lambda_k^{-1})t_{ik}^T \qquad [7]$$

where t_{ik} is PC scores for i^{th} sample spectrum using component k, and λ_k^{-1} is the diagonal matrix of normalized eigenvalues of the covariance matrix for component k. λ_k^{-1} is used as a weighting factor for the PCs contributing with different variances. Therefore, Hotelling's T^2 gives an indication of extreme values within the PCA model. On the other hand, Q-residuals is a measure of variance which is not captured by the PCA model (lack of model fit statistics) and is defined by[149]:

$$Q_{ik} = \sum (x_i - t_{ik} P_k^T)^2 \quad [8]$$

where x_i is the sample spectrum, Q_{ik} is the sum of squared reconstruction error for i^{th} sample spectrum using component k and P_k is the PC loading vectors. For both Hotelling's T^2 and Q-residuals, the normalized 99% CI was utilized as upper thresholds to intercept anomalous Raman spectra (e.g., non-contact measurements, spectra from blood, etc.) Accordingly, the Hotelling's T^2 and Q-residuals are two independent parameters providing quantitative information about the model fit. Using these parameters as indicators of spectrum quality (i.e., probe contact mode, confounding factors, white light interference, etc.), auditory feedback has been integrated into the on-line Raman diagnostic system, facilitating real-time probe handling advice and spectroscopic screening for clinicians during clinical endoscopy procedures.

6.2.2 On-line probabilistic diagnostics

Subsequent to outlier analysis of tissue Raman spectra quality, those qualified Raman spectra are immediately fed to probabilistic models for on-line *in vivo* diagnostics and pathology prediction. The GUI can instantly switch among different models including PLS-DA, PCA-LDA, ACO-LDA, CART, SVM, adaptive boosting (Ada-Boost), etc. for prospective classification at clinical endoscopic procedures. In this study, as an example, probabilistic PLS-DA was employed for gastric cancer diagnosis. The developed system also supports binary classification, one-against-all and one-against-one multiclass (i.e., benign, dysplasia, and cancer) probabilistic PLS-DA discriminatory analysis to predict the specific tissue pathologies.

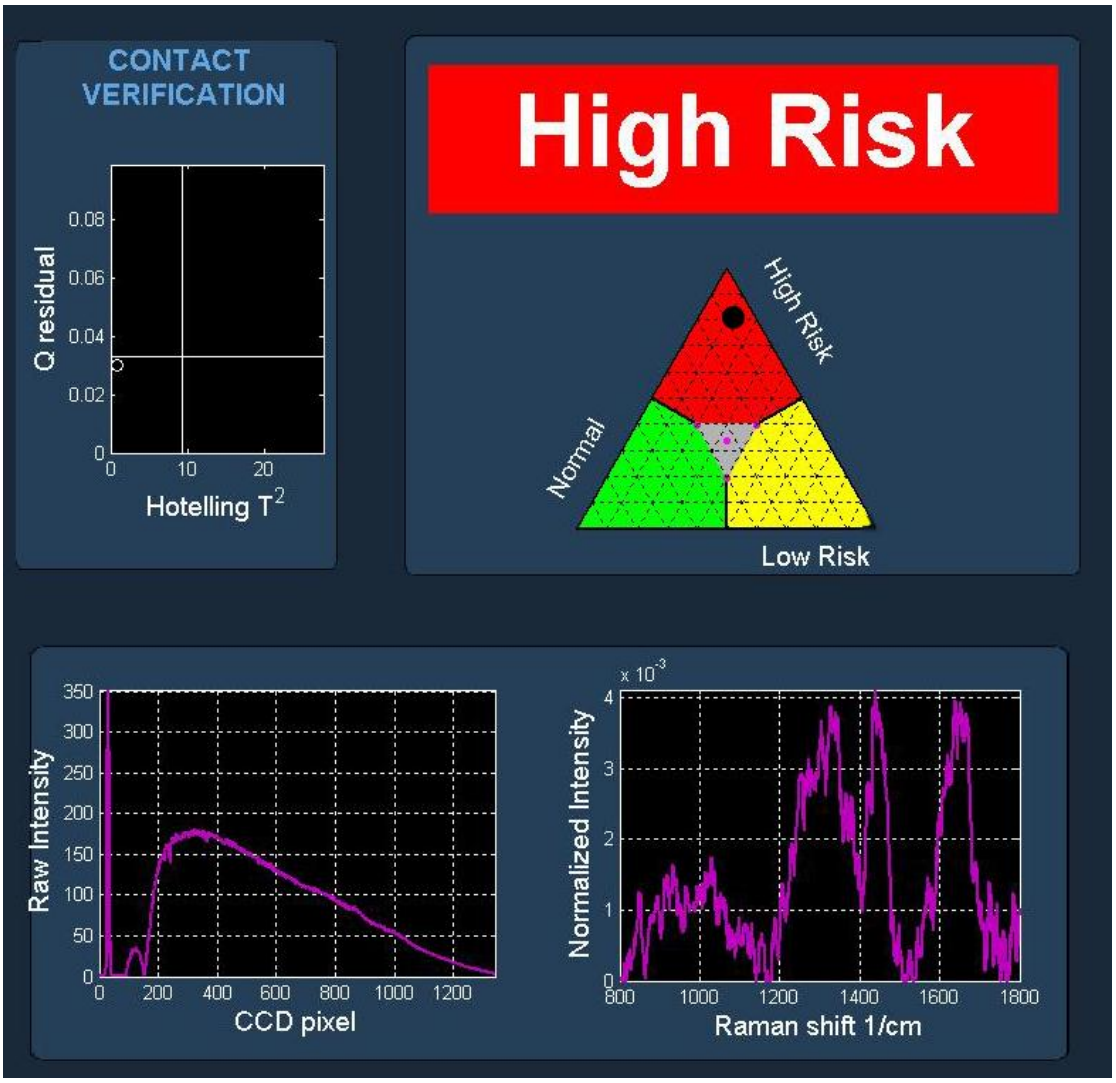


Figure 6.2 On-line biomedical Raman spectroscopic diagnostic framework developed.

6.1 Results

Figure 6.2 shows the developed GUI for on-line biomedical spectroscopic processing and diagnostics. We tested the developed on-line framework in the stomach that represents one of the most challenging organs presenting with many confounding factors (i.e., gastric juice, food debris, bleeding, exudates, etc.) for spectroscopic diagnosis. The *in vivo* mean Raman spectra acquired from 305 gastric patients (normal (n = 2465) and cancer (n = 283)) for algorithms development are shown in Figure 6.3. A detailed analysis and discussion of the Raman spectral characteristics of carcinogenesis in the gastric was presented in Chapter 5.

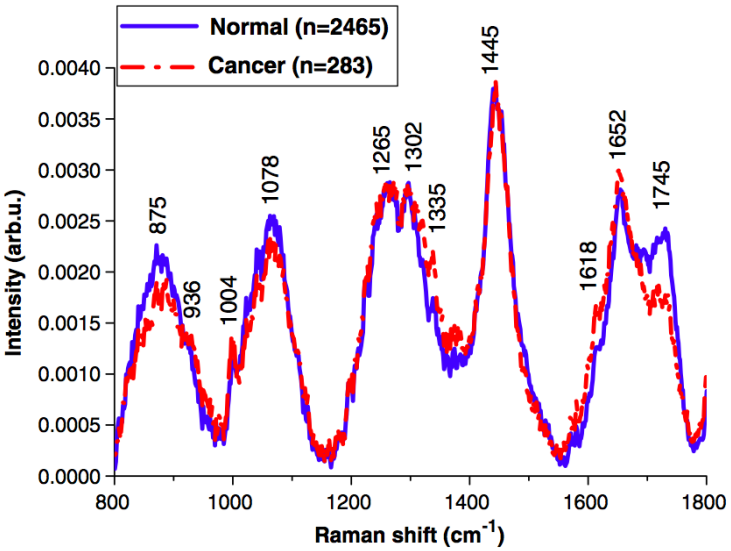


Figure 6.3 *In vivo* mean Raman spectra of normal (n = 2465) and cancer (n = 283) gastric tissue acquired from 305 gastric patients.

The automatic outlier detection was realized for predictive on-line analysis using PCA with Hotelling's T^2 and Q-residuals statistics (99% CI). A two-component PCA model was rendered that included the largest tissue spectral variations. These selected significant PCs ($p < 0.0001$) accounted for maximum variance of 38.71% (PC1: 30.33%, PC2: 8.38%) of the total variability in the dataset (n = 2748 Raman spectra), and the corresponding PC loadings are shown in Figure 6.4 .

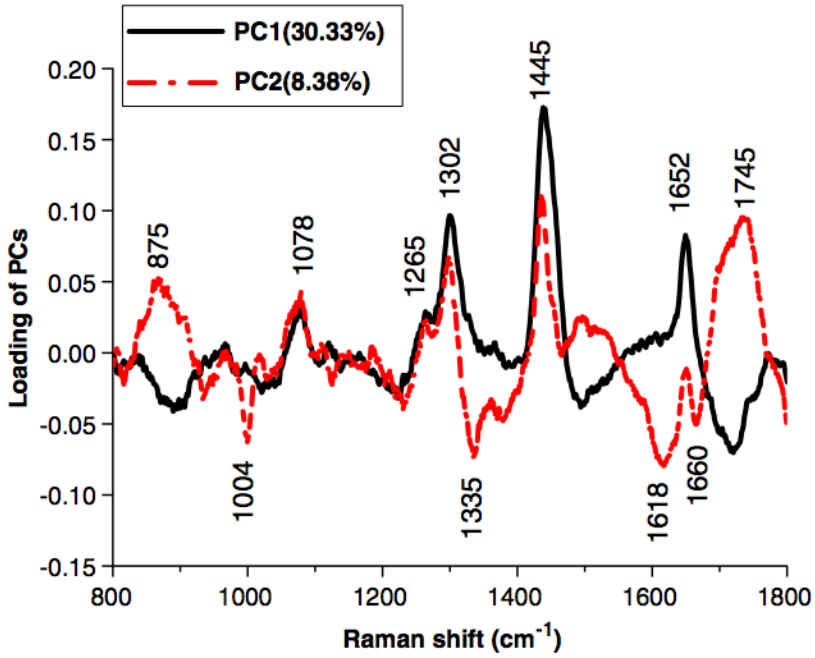


Figure 6.4 The principal component (PC) loadings calculated from the spectral training database of 2748 (2465 normal and 283 cancer) Raman spectra of gastric tissue *in vivo*. The significant PCs capture a total variance of 38.71% (PC1, 30.33%; PC2, 8.38%).

Figure 6.5 shows the score scatter plots (i.e., PC1 versus PC2) for the normal (n = 2465) and cancer tissue spectra (n = 283), exemplifying the capability of PC scores for separating the cancer spectra from normal. The 99% CI of Hotelling's T^2 and Q residuals were accordingly calculated from the training dataset and fixed as a threshold for prospective on-line spectral validation.

We then rendered probabilistic PLS-DA models for prediction of gastric cancer. The training database was randomly resampled multiple times (n = 10) into learning (80%) and test (20%) sets. The generated PLS-DA models provided a predictive accuracy of 85.6% (95% CI: 82.9% to 88.2%) (sensitivity of 80.5% (95% CI: 71.4% to 89.6%) and specificity of 86.2% (95% CI: 83.6% to 88.7%)) for gastric cancer diagnosis. We subsequently tested the online framework (i.e., outlier-detection

as well as probabilistic PLS-DA) in 10 prospective gastric patients. PC score scatter plots (i.e., PC1 versus PC2) for the prospective normal (n = 45) and cancer (n = 30) tissue spectra are also shown in Figure 6.5.

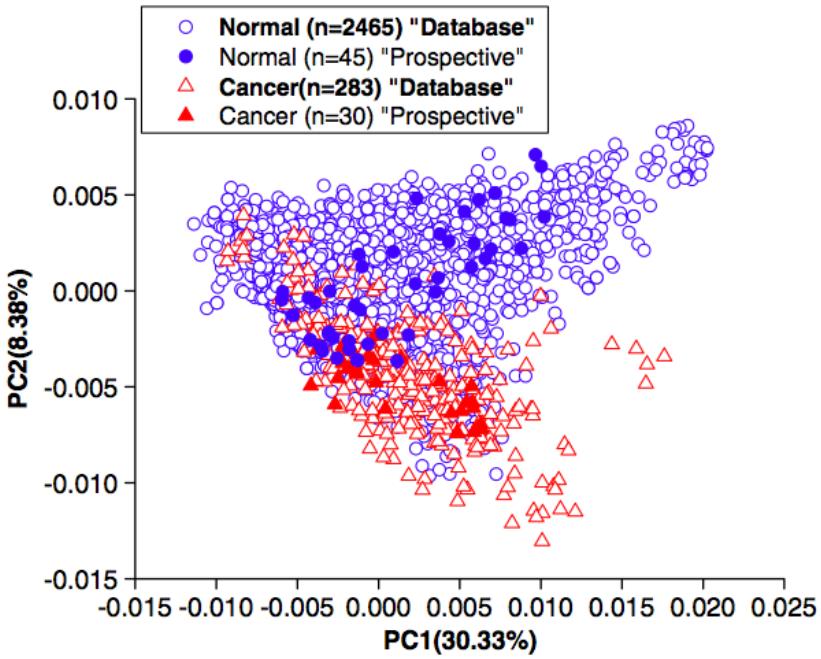


Figure 6.5 Scatter plots of the two diagnostically significant PC scores (PC1 versus PC2) with a total variance of 38.71% (PCl, 30.33%; PC2, 8.38%) for 2465 normal (training database) and 283 cancer (training database) gastric tissue Raman spectra, together with the PC scores of prospective 45 normal and 30 gastric cancer Raman spectra.

Figure 6.6 shows the prospective scatter plot of the Hotelling's T^2 (38.71%) and Q-residuals (61.29%) with the 99% CI boundaries for 105 spectra acquired from the prospective gastric patients. It is observed that a large number of noncontact spectra lie outside the 99% CI and are therefore discarded in real-time. The verified tissue Raman spectra largely fall inside the 99% CI of T^2 and Q residuals, demonstrating that this on-line data analysis provides a rapid and highly efficient means of real-time validation of biomedical tissue spectra. The prospectively acquired spectra verified by the on-line outlier analysis are further fed to probabilistic PLS-DA for instant disease

prediction, achieving a diagnostic accuracy of 80.0% (60⁄75) (sensitivity of 90.0% (27⁄30), and specificity of 73.3% (33⁄45)) for gastric cancer detection (Figure 6.7), as confirmed by histopathological examination.

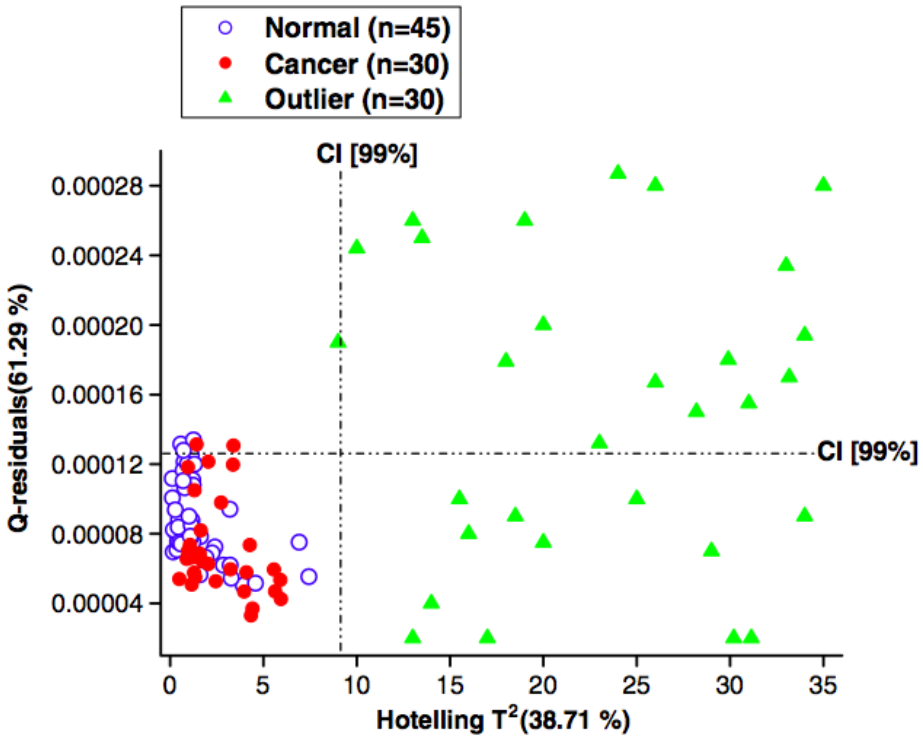


Figure 6.6 Hotelling's T^2 versus Q-residuals for 105 Raman spectra (45 normal, 30 cancer, 30 outlier) acquired from 10 prospective gastric samples. The outlier (n = 30 green triangles) shows higher T^2 and higher Q values than the valid spectra. The dotted line represents the 99% confidence interval (CI) verifying whether the prospective Raman spectra are within the common tissue variations of the principal component analysis (PCA) model.

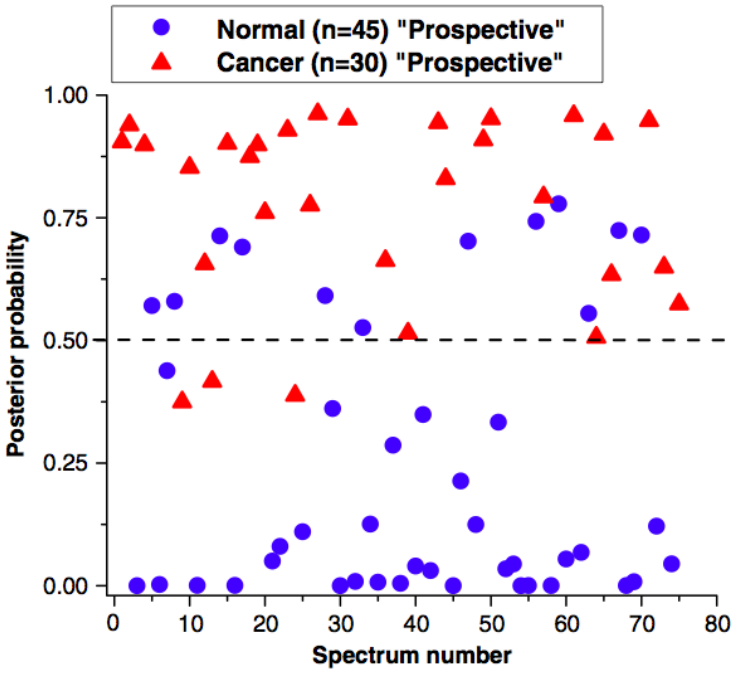


Figure 6.7 Scatter plot of the posterior probability values belonging to prospective normal (n = 45) and cancer (n = 30) gastric tissue based on PLS-DA modeling. The separate dotted line gives a diagnostic sensitivity of 90.0% (27⁄30) and specificity of 73.3% (33⁄45) for separating cancer from normal gastric tissue *in vivo*.

The ROC curves were further generated to evaluate the group separations. Figure 6.8 shows the mean of the ROC curves computed from each random splitting of the spectral database for retrospective prediction as well as the ROC calculated for the prospective dataset prediction. The integration areas under the ROC curves generated for the retrospective and prospective datasets are 0.90 and 0.92, respectively, illustrating the robustness of the PLS-DA algorithm for prospective gastric cancer diagnosis *in vivo*.

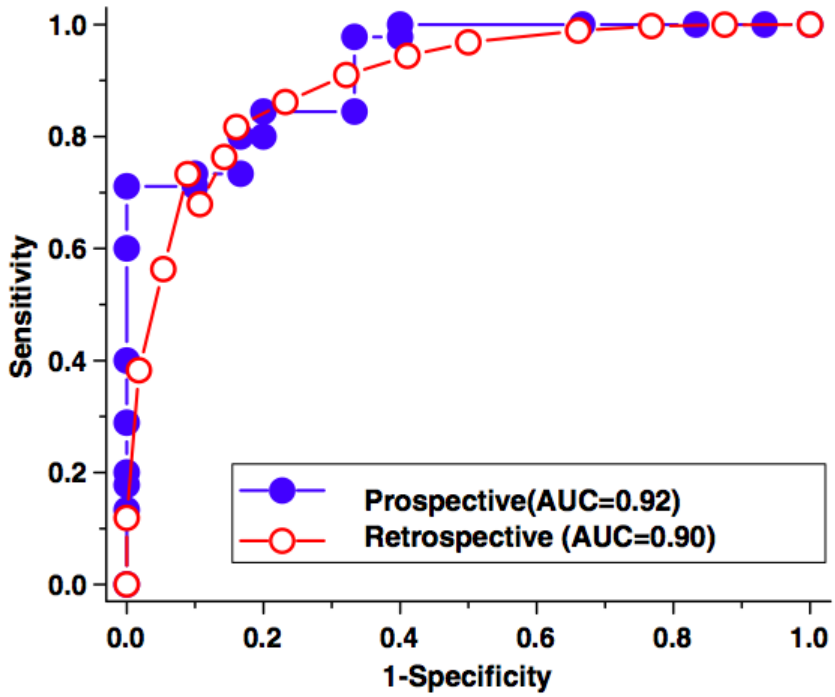


Figure 6.8 The ROC curves computed from the spectral database for retrospective prediction as well as ROC curve for prospective prediction of normal and cancer gastric tissue. The integrated areas under the ROC curves are 0.90 and 0.92 for the retrospective and prospective datasets, respectively, illustrating the efficacy of Raman endoscopy together with PLS-DA algorithms for *in vivo* gastric cancer diagnosis.

The total processing time for the framework was 0.13 s. The processing time for each step of the flow chart (Figure 6.1) are given in Table 7. Free-running optical diagnosis and processing time of < 0.13 s can be achieved, which is critical for realizing real-time *in vivo* tissue diagnostics at endoscopy. Overall, this work demonstrates for the first time that the on-line biomedical diagnostic framework can move the biomedical Raman spectroscopy into real-time, on-line cancer detection and diagnosis during routine clinical endoscopic examination.

Analyses	Computational time (milliseconds)
Cosmic ray rejection	0.5
Laser response time	10
Preprocessing	40
Outlier detection	10
Probabilistic PLS-DA prediction	70
Total computation time	100 to 130

Table 7 Average processing time for on-line biomedical Raman spectroscopic framework on a personal computer with a 64-bit I7 quad-core 4GB memory.

6.2 Discussion

We have presented an on-line framework for biomedical Raman spectroscopy that integrates preprocessing, automatic outlier detection, and deploys instant change of diagnostic models among different organs in the upper GI tract during gastroscopy. We have introduced automated outlier detection for spectra verification in endoscopic applications. PCA with related Hotelling's T^2 and Q-residuals is a high-level metrics for outlier detection, and we found that it is highly efficient and reliable in our on-line diagnostic platform at endoscopic diagnosis. The PCA model was generated from the database accounting for the majority of common tissue variations (38.71% of total variance, PC1: 30.33%; PC2: 8.38%) (Figure 6.4). The rendered PCA model was further tested prospectively on 10 gastric patients for prospective automatic outlier detection during *in vivo* on-line Raman spectroscopic screening (Figure 6.6). Of the total 105 gastric spectra acquired from the prospective gastric patients, 75 spectra that fit the developed PCA model well (T^2 and Q lies within 99% CI (blue and red circles)) were eligible for further classification. The remaining spectra (green triangles) from noncontact measurements, food debris or blood contaminations show unusual disturbances, which were poorly reconstructed by the PCA model. Large Q-

residuals indicated the presence of spectral variations unexplainable by the model due to many confounding factors (e.g., noncontact mode, contaminants, or unknown spectral variations) frequently occurring in *in vivo* clinical measurements. One noted that tissue Raman spectra presenting with Q-residuals vaguely outside the CI ((n = 6) were accepted following off-line analysis because the newly introduced spectral variations could be due to tumor invasion, necrosis, tumor subtype, etc. Thus following on-line analysis, it is advisable to review the spectra in off-line mode to assess possible new spectral variations. This study thoroughly shows the necessity of utilizing automated outlier diagnostics for tissue Raman spectral quality verification in real-time endoscopic applications.

Subsequent to on-line outlier analysis, the spectra were fed to probabilistic PLS-DA algorithms for disease prediction. The PLS-DA modeling provided the predictive accuracy of 80.0% (60⁄75) (sensitivity of 90.0% (27⁄30) and specificity of 73.3% (33⁄45)) (Figure 6.7) for cancer diagnosis on 10 prospective gastric patients, suggesting that Raman endoscopy with the integration of on-line diagnostic framework can be a diagnostic screening tool for real-time *in vivo* gastric cancer identification. These probabilistic diagnostic outcomes were presented in real-time to the endoscopists within 0.13 s, which is a clinically acceptable time at endoscopy. With the immediate auditory diagnostic feedback from the GUI developed, the endoscopists can now perform point-wise scanning for targeted biopsies of the high-risk tissue sites. The movie (Supplementary information: http://biomedicaloptics.spiedigitallibrary.org/article.aspx?articleid=1352558) shows the on-line Raman endoscopic diagnostic procedure for real-time *in vivo* detection of gastric cancer during clinical endoscopic examination.

6.1 Conclusion

In this chapter, we have developed an on-line biomedical Raman spectroscopic framework to translate the cumbersome processing and multivariate/chemometric analysis into an easy-to-use GUI interface for real-time, *in vivo* prospective diagnosis of malignancies in the GI tract. The efficacy of the on-line diagnostic framework integrated with the multimodal image-guided clinical Raman spectroscopy was proven for the first time on the prospective gastric patients.

Chapter 7. Fiber-optic confocal Raman spectroscopy for real-time *in vivo* diagnosis of dysplasia in Barrett's esophagus

In this chapter we apply the developed on-line Raman technique prospectively in the BE esophagus. BE is a condition that poses high risk of developing dysplasia leading to esophageal adenocarcinoma. Early detection of dysplasia in BE is critical but remains endoscopically challenging. The current diagnostic approaches for detecting BE relies on random biopsy sampling guidelines, resulting in a vast number of negative biopsies. To improve the sensitivity of Raman technique for precancerous lesions, we develop a novel fiber-optic confocal Raman spectroscopy technique. We acquire a spectral reference database of 373 patients and demonstrate the prospective diagnostics in 77 patients *in vivo* during clinical endoscopy. High-quality epithelial Raman spectra can be acquired from esophagus *in vivo* and processed in real-time (<0.5 second). Confocal Raman spectra show significantly spectral differences ($p < 0.001$) among different esophageal tissue types (i.e., columnar lined, non-dysplastic BE, and dysplastic BE). Fiber-optic confocal Raman spectroscopy provides a diagnostic sensitivity of 87.0% (67/77) and a specificity of 84.7% (610/720) for *in vivo* detection of dysplasia in 77 patients prospectively based on endogenous biomolecular signatures. This chapter demonstrates for the first time that the real-time fiber-optic confocal Raman spectroscopy can be used to prospectively detect dysplasia in BE *in vivo* at the molecular level during clinical endoscopy.

7.1 Introduction

BE is a metaplastic precursor of EAC. Since the late 1970s, EAC has had a substantial rise in incidence rates (>350%), and is growing more rapidly than other cancers in the developed countries, despite of extensive efforts for prevention, including periodic surveillance of high-risk BE patients.[13, 14] The incidence of EAC is continuously increasing and the five-year survival rates remain low. Given the poor therapeutic response of symptomatic EAC, early identification of high-risk lesions (i.e., dysplasia) together with therapeutic interventions are currently the most critical measures to improving survival rates of BE patients.[13, 14] However, dysplastic lesions or grossly inconspicuous cancers can be endoscopically indistinguishable from the surrounding metaplastic tissue. This is because conventional endoscopy heavily relies on visual assessment of structural and morphological features of tissue surface, resulting in poor diagnostic accuracy. Existing diagnostic guidelines recommend extensive biopsy samplings (typically four-quadrant samplings) at every 1-2 cm interval along suspicious Barrett's segments during endoscopic inspections of BE patients. This approach produces a vast number of negative biopsies, and is clinically labor intensive and a burden to the patients. Since only a minute amount of the mucosa is sampled (as little as 5%), tissue biopsies do not accurately characterize BE segments. Foci of dysplasia in a background of IM are frequently overlooked even when the biopsies are diligently performed by experienced endoscopists using extensive four-quadrant biopsy protocols. Taken into account the enormous rise in incidence of EAC and the existing clinical challenges, the need for new advanced endoscopic modalities has never been greater. The objective targeting of high-risk tissue areas with a minimally invasive technique could greatly reduce random biopsy sampling errors as well as health care expenses and burden on the patients. In 2004,

Motz et al. presented a 2 mm diameter sapphire ball lens coupled fiber-optic probe.[66] In this report, we present the first *in vivo* application of a fiber-optic confocal Raman spectroscopy technique for real-time prospective diagnosis of dysplasia in BE at endoscopy.

7.2 Materials and methods

7.2.1 Fiber-optic confocal Raman instrumentation

The developed fiber-optic confocal Raman spectroscopic system (Figure 7.1) consists of a NIR diode laser (λ_{ex} = 785 nm), a high-throughput transmissive imaging spectrograph equipped with a liquid nitrogen-cooled, NIR-optimized CCD camera and a specially designed 1.8-mm (in outer diameter) fiber-optic confocal Raman probe.[125] The system acquires Raman spectra in the range of 800-1800 cm^{-1} with a spectral resolution of ~9 cm^{-1}. The developed fiber-optic confocal Raman endoscopic probe is used for both laser light delivery and *in vivo* tissue Raman signal collection.[6] The 1.8 mm confocal Raman endoscopic probe comprises 9 × 200 μm filter-coated collection fibers (NA = 0.22) surrounding the central light delivery fiber (200 μm in diameter, NA = 0.22). A miniature 1.0 mm sapphire ball lens (NA = 1.78) is coupled to the fiber tip of the confocal probe to tightly focus the excitation light onto tissue, enabling the effective Raman spectrum collection from the epithelial lining (< 200 μm).[125] The fiber-optic confocal Raman probe can be inserted into the instrument channel of conventional endoscopes and placed in gentle contact with the epithelium for *in vivo* tissue characterization and diagnosis.

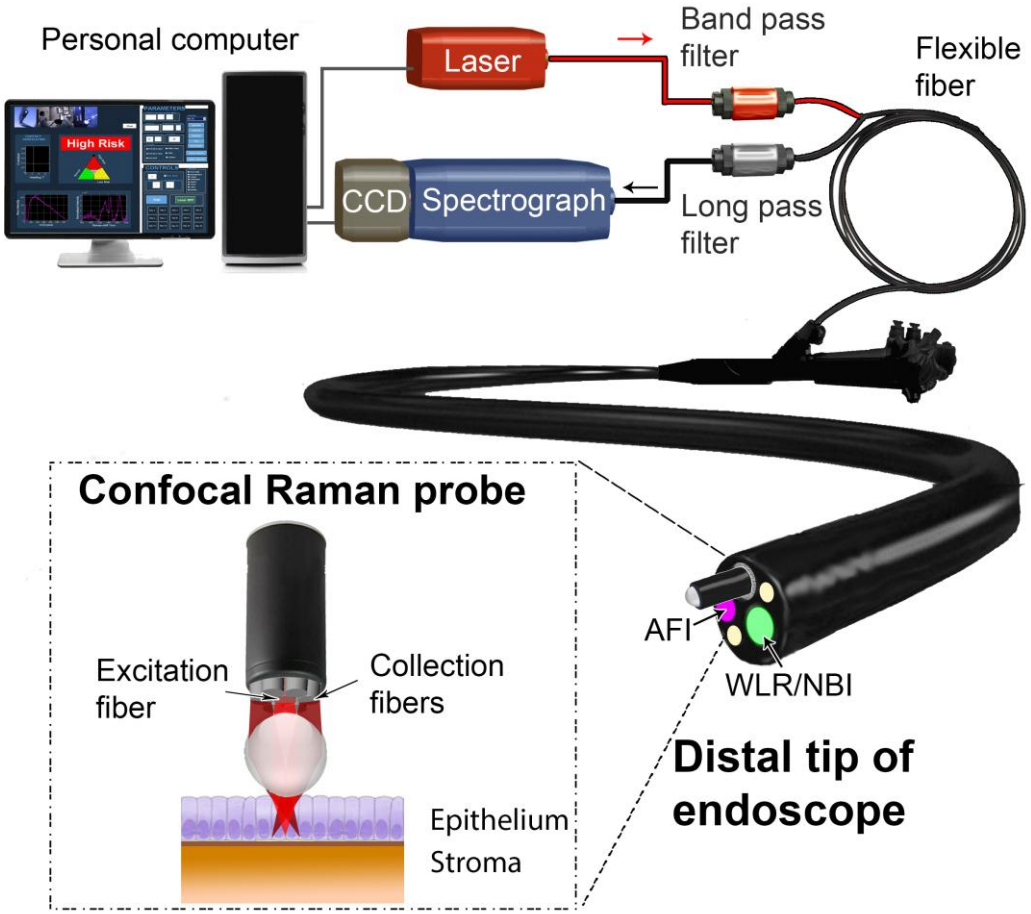


Figure 7.1 The rapid fiber-optic confocal Raman spectroscopy system developed for *in vivo* tissue diagnosis and characterization at endoscopy.

The depth-selectivity of this probe with different bevelled angles of the collection fibers gives the instrument confocal properties (Figure 7.2) which offers compelling experimental advantages,[125] including (i) fiber-optic confocal Raman spectroscopy selectively targets the epithelial lining associated with early onset of Barrett's carcinogenesis which is superior to conventional volume-type fiber-optic Raman probes that interrogate a larger tissue volume; (ii) the shallower tissue interrogation ability of confocal Raman technique suppress tissue autofluorescence contribution and interferences from deeper tissue layers (e.g., stroma),[125] (iii) confocal Raman technique enables more reproducible measurements (i.e., Raman signal is only detected at accurate tissue contact).[125]

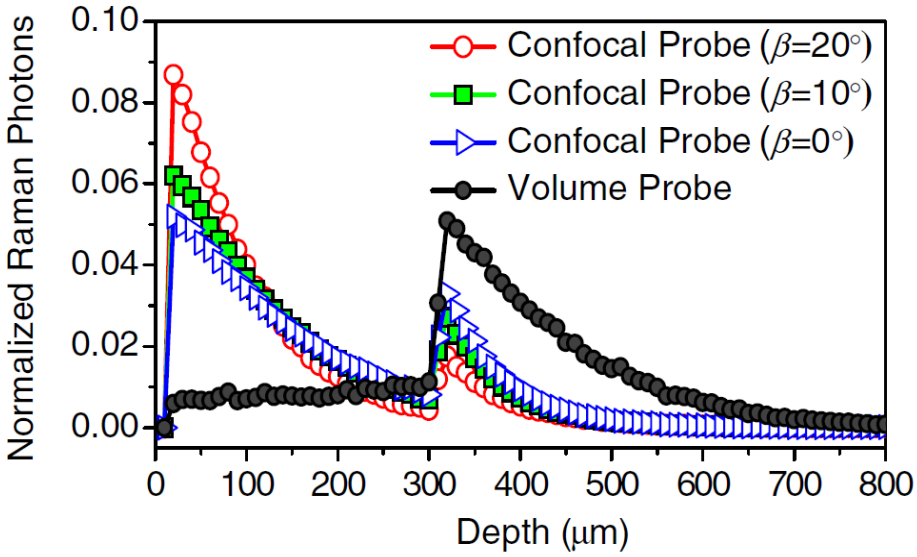


Figure 7.2 Monte Carlo simulations: Normalized Raman photons collected as a function of tissue depths using different beveled fiber confocal Raman probes (β = 0°, 10°, and 20°; gap d = 0 μm) as well as the Raman photons collected using a volume-type Raman probe in a two-layer tissue model.[152]

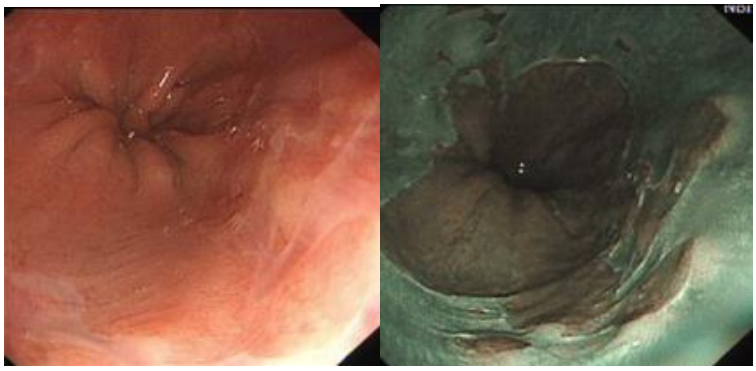


Figure 7.3 WLR and NBI image of a patient with Barrett's esophagus. The tongues of salmon appearing mucosa extending into the esophagus are associated with increased risk of developing esophageal adenocarcinoma. The images were obtained from Endoscopy Center, National University Hospital of Singapore (NUHS) with permission.

7.2.1 Data processing

The data processing has been reported in detail in Section 4.2.2

7.2.2 Study protocol

The patient recruitment and ethical protocol has been reported in detail in Chapter 3. Biopsies are classified into the categories: Columnar-lined epithelium (CLE), nondysplastic BE defined as the presence of goblet cells, BE indeterminate for dysplasia, BE positive for low-grade dysplasia, and BE positive for HGD. The histopathology results (the gold standard) are compared with Raman measurements to determine the diagnostic performance of the confocal Raman technique for identifying dysplasia in BE.

7.3 Results and discussion

Figure 7.4 shows the mean *in vivo* confocal Raman spectra measured from the database of 373 patients presenting with different tissue types (CLE, n=907; nondysplastic BE, n=318; BE positive for HGD, n=177) as confirmed by histopathologic characterization. Prominent tissue Raman peaks are observed at 936 cm^{-1} (υ(C-C) proteins), 1004 cm^{-1} (υ_s(C-C) ring breathing of phenylalanine), 1078 cm^{-1} (υ(C-C) of lipids), 1265 cm^{-1} (amide III υ(C-N) and δ(N-H) of proteins), 1302 cm^{-1} (CH_2 twisting and wagging of lipids), 1445 cm^{-1} (δ(CH_2) deformation of proteins and lipids), 1618 cm^{-1} (ν(C=C) of porphyrins), 1655 cm^{-1} (amide I υ(C=O) of proteins), and 1745 cm^{-1} (υ(C=O) of lipids) (Table 1).[25, 94, 125] Raman spectral differences (e.g., peak intensity, shifting, and band broadening) can be discerned among different tissue types ($p < 0.001$, unpaired 2-sided Student's t-test). While histopathology (Figure 7.5 R) identifies the presence of intestinal mucosa marked by goblet cells in nondysplastic BE as well as progressive architectural and cytologic atypia in dysplastic BE, the fiberoptic confocal Raman spectroscopy uncovers the biochemical and biomolecular changes occurring in the epithelium accompanying Barrett's carcinogenesis. For

instance, the Barrett's tissue areas with pathologically confirmed HGD showed significant increased Raman intensities at 1004 cm^{-1} (phenylalanine) and 1655 cm^{-1} (amide I of proteins; $p < 0.001$, unpaired 2-sided Student's t-test) compared with nondysplastic BE owing to up-regulated proteins content associated with an increased proliferation of epithelial cells in dysplastic BE.[91] The distinct Raman peaks such as 1335 and 1576 cm^{-1} (Table 1) of DNA are also closely linked with nuclei abnormalities in dysplastic tissue and cells (Figure 7.5 R). One notes that the Raman biomolecular signature of CLE is distinct, but the nondysplastic BE bears a resemblance to that of dysplastic BE to a high degree, suggesting that transformation to IM phenotype is accompanied with prominent molecular abnormalities, which could be a key event in Barrett's carcinogenesis. Overall, the highly specific Raman molecular signatures observed reflect a multitude of endogenous optical biomarkers (e.g., oncoproteins, DNA content, mucin expression) in epithelial tissue.[91] The correlation of the epithelial Raman spectral signatures with histopathology and histochemistry can deepen the understanding of Barrett's onset and progression *in situ* at the molecular level.

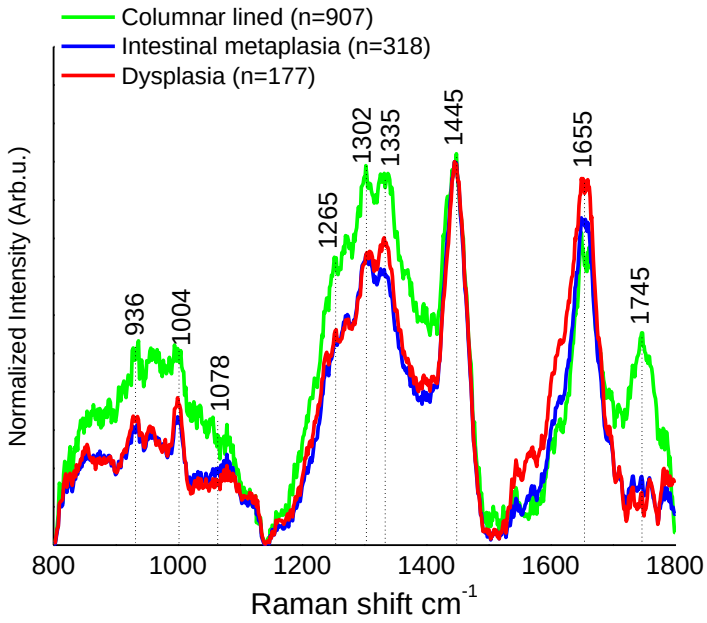


Figure 7.4 The mean *in vivo* confocal Raman spectra of columnar-lined epithelium (n=907), intestinal metaplasia (n=318), dysplasia (n=177) acquired from 373 patients during clinical endoscopic examination. Each Raman spectrum is acquired within 0.1 to 0.5 sec. The spectra have been normalized to the Raman peak at 1445 cm^{-1} for comparison purpose.

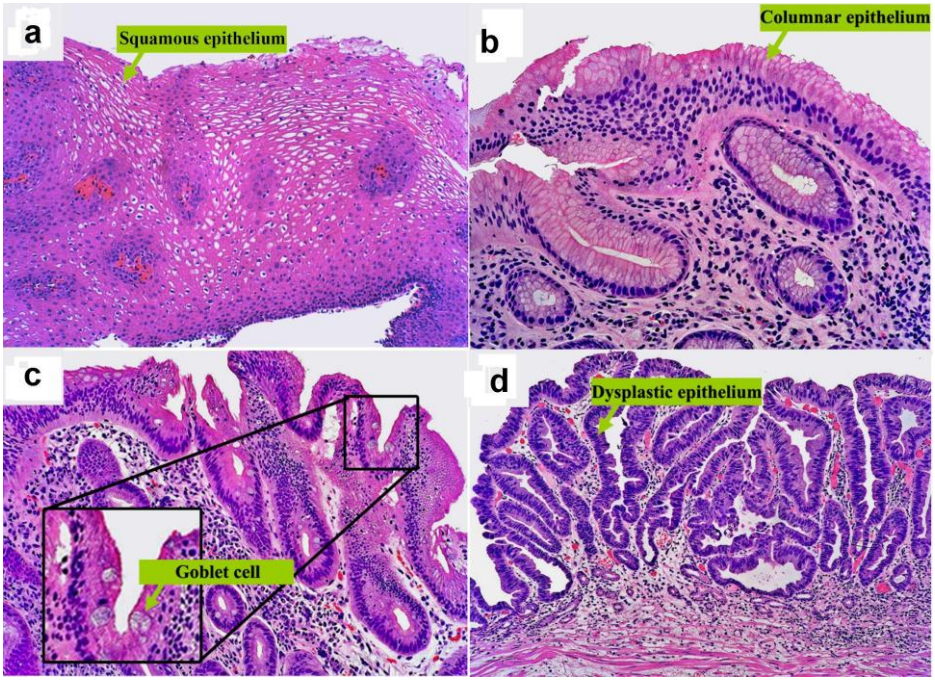


Figure 7.5 Representative hematoxylin and eosin (H&E)-stained histopathologic slides corresponding to different esophageal tissue types measured. (a) Squamous-lined epithelium, and (b) Columnar lined esophagus with absence of goblet cells (original magnification, 200×). (c) Nondysplastic Barrett's esophagus, where the normal stratified squamous epithelium is replaced by intestinal metaplastic (IM) epithelium containing goblet cells (original magnification, 200×). (d) High-grade dysplasia in Barrett's esophagus showing both architectural and cytologic atypia as well as crowded crypts with branching and papillary formation, cytologic pleomorphism and loss of polarity (original magnification, 100×).

Capitalizing on the rich epithelial Raman signatures acquired from 373 patients, we have applied the confocal Raman spectroscopy technique together with multivariate diagnostic algorithms developed for *in vivo* prediction and diagnosis of prospective patients with suspicious BE. Figure 7.6 shows the 2-dimensional ternary plot of the posterior probabilities using trichotomous probabilistic PLS-DA on 77 prospective patients belonging to (i) CLE (n=597), (ii) nondysplastic BE (n=123), and (iii) HGD BE (n=77), respectively, using confocal Raman spectroscopy technique. The corresponding receiver operating characteristic (ROC) curves of dichotomous discriminations among CLE versus nondysplastic BE + HGD BE, nondysplastic BE versus CLE + HGD BE, and HGD BE versus CLE + nondysplastic BE are shown in Figure 7.7, with the areas under ROC curves of being 0.88, 0.84, and 0.90, respectively, for different tissue classification. The ROC shows that fiber-optic confocal Raman spectroscopy provides a diagnostic sensitivity of 87.0% (67/77), and a specificity of 84.7% (610/720) for *in vivo* detection of HGD in BE. In the light of these promising prospective results, fiber-optic confocal Raman spectroscopy represents a potent optical diagnostic means to revealing the detailed endogenous biomolecular information of epithelial tissue, thereby enabling *in vivo* objective diagnosis of dysplasia in BE patients.

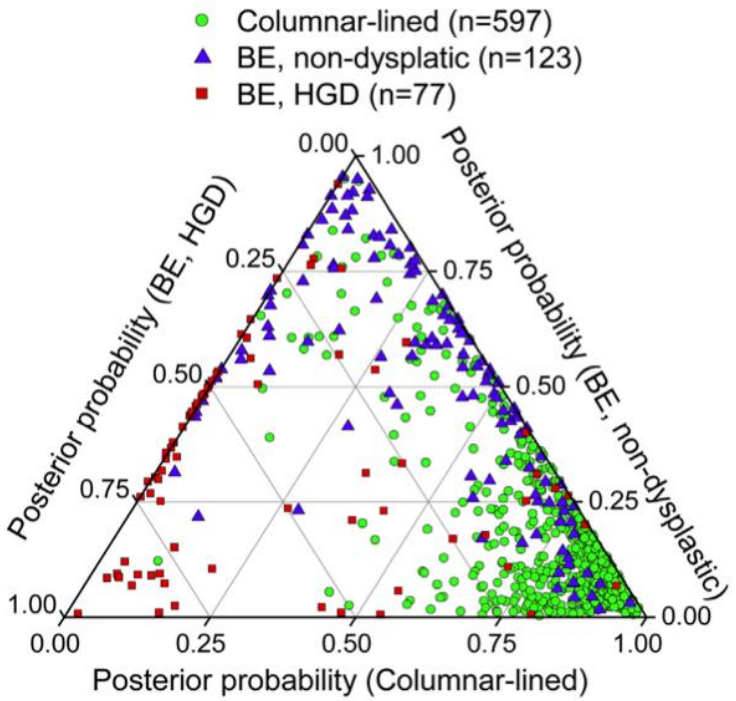


Figure 7.6 Two-dimensional ternary plot of the posterior probabilities of prospective 77 patients belonging to columnar-lined epithelium (CLE; n=597), nondysplastic Barrett's esophagus (BE; n = 123), and high-grade dysplastic BE (n = 77), respectively, using confocal Raman spectroscopy technique.

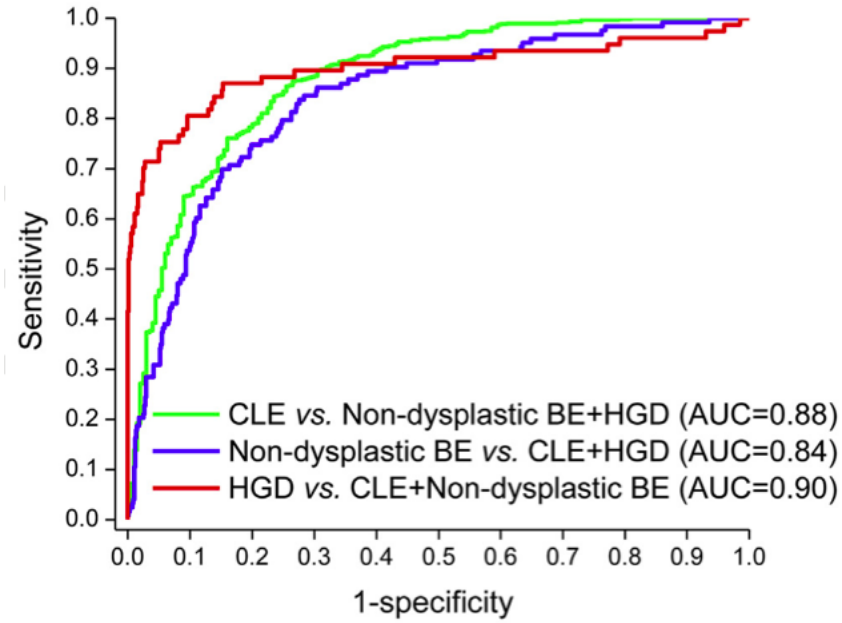


Figure 7.7 Receiver operating characteristic (ROC) curves of dichotomous discriminations among CLE versus nondysplastic BE + HGD BE, nondysplastic BE versus CLE + HGD BE, and HGD BE versus CLE + nondysplastic BE, with the areas under ROC curves (AUC) of being 0.88, 0.84, and 0.90, respectively.

7.4 Video Descriptions

We demonstrated the clinical utility of fiber-optic confocal Raman spectroscopy for guiding the biopsies of suspicious BE at endoscopy (Supplemental Video, http://www.gastrojournal.org/article/S0016-5085%2813%2901585-0/addOns). A 54-year-old Chinese man was scoped on the upper GI with conventional white-light reflectance (WLR) endoscopy and subsequently narrow-band imaging (NBI). Using WLR endoscopy, the suspicious Barrett's segment appeared salmon-colored with a tongue of columnar lined epithelium or BE extending approximately 2 cm into the esophagus. Under NBI, the adjacent normal squamous epithelium seemed greenish in color. There was no evidence of morphologic/structural or microvascular changes in the suspicious Barrett's segment to indicate the presence of dysplasia. After preliminary visualization under WLR/NBI, fiber-optic confocal Raman scans were performed on the suspicious Barrett's tongues. Confocal Raman spectral diagnosis was given in real-time (<0.2-0.5 sec) with auditory feedback, which identified the suspicious segment as a "low-risk" lesion corresponding to nondysplastic BE. Subsequent biopsies and histopathologic examinations confirmed that the measured tissue areas indeed were nondysplastic in nature, but with moderate to extensive presence of goblet cells.

This study demonstrates for the first time that fiberoptic confocal Raman spectroscopy can be used to target dysplasia in BE patients in real-time. By enabling functional and biomolecular assessment, fiberoptic confocal Raman spectroscopy constitutes a new endoscopic modality for in vivo tissue diagnosis and characterization. The introduction of the objective diagnostic modality such as this could have a major impact on current endoscopic practice and clinical decision making. The use of fiberoptic confocal Raman spectroscopy may include many

clinical scenarios that require histopathologic biopsy samplings in screening of patients at higher risk of developing EAC. Ultimately, the targeting of dysplastic lesions and the discrimination between LGD and HGD, the clinical hallmarks in Barrett's carcinogenesis, will possibly be the key metric for defining the true clinical value of fiberoptic confocal Raman spectroscopy. Because dysplastic foci are prone to biopsy sampling errors, fiberoptic confocal Raman spectroscopy could be used to accurately pinpoint these high-risk tissue areas in vivo in real-time as well as examine those lesions that are indefinite for dysplasia or presenting with extensive inflammation where histopathologic manifestations are insufficient for clinical decision making. It may even be possible to detect field carcinogenesis or prognostic information in patients occurring before histopathologic manifestations. Fiberoptic confocal Raman spectroscopy could therefore be used to objectively stratify those patients who have higher risk of developing EAC and prompt increased frequency of surveillance or therapeutic intervention for the individual patients. These facets of confocal Raman technology certainly warrant further investigations. Preemptive approaches or interventional procedures for HGD or intramucosal cancers are other scenarios where fiberoptic confocal Raman spectroscopy could have a potential key role to play. Confocal Raman endoscopy offers the gastroenterologist a novel objective tool, which allows real-time assessment of tissue pathology in situ, and therefore could be used to accurately define theresection margins of macroscopically inconspicuous dysplasia or cancers in BE patients undergoing surgical intervention (eg, endoscopic mucosal resection, endoscopic submucosal dissection, ablation). This could enable complete excision and subsequent margin assessment, so that malignant progression can be efficiently prevented. Today, the necessity for histopathologic characterization of biopsy specimens hinders diagnosis and therapeutic intervention

during the same endoscopic procedure. Confocal Raman technology opens up the possibility for combining final diagnosis and therapeutic eradication (eg, endoscopic mucosal resection, endoscopic submucosal dissection) in a single procedure, thereby effectively reducing medical cost and burden associated with multiple endoscopic procedures.

7.5 Conclusion

We have demonstrated in this chapter that real-time fiber-optic confocal Raman spectroscopy can be performed prospectively in screening of the patients with suspicious BE *in vivo*. Fiber-optic confocal Raman spectroscopy uncovers the functional and biomolecular changes occurring in the epithelium during Barrett's carcinogenesis. Capitalizing on the rich epithelial Raman signatures acquired, fiber-optic confocal Raman spectroscopy can be used to objectively target dysplasia in BE, illustrating its utility for improving *in vivo* precancer diagnosis and tissue characterization of BE at the molecular level during gastrointestinal endoscopy.

Chapter 8. Development of a multiplexed fingerprint and high wavenumber Raman spectroscopy technique for real-time *in vivo* tissue Raman measurements at endoscopy

In this chapter, we develop a multiplexed Raman spectroscopy technique using a single laser light together with a volume phase holographic (VPH) grating that simultaneously acquires both FP and HW tissue Raman spectra at endoscopy. We utilize a customized VPH dual-transmission grating, which disperses the incident Raman scattered light vertically onto two separate segments (i.e., 150 to 1950 cm^{-1}; 1750 to 3600 cm^{-1}) of a CCD camera. We demonstrate that the multiplexed Raman technique can acquire high quality *in vivo* tissue Raman spectra ranging from 800 to 3600 cm^{-1} within 1.0 s with a spectral resolution of ~3 to 6 cm^{-1} during clinical endoscopy. The rapid multiplexed Raman spectroscopy technique covering both FP and HW ranges developed in this work has potential for improving *in vivo* tissue diagnosis and characterization at endoscopy

8.1 Introduction

To date, most NIR Raman studies have been centered on the FP range (i.e., 800 to 1800 cm^{-1}) owing to the wealth of specific biomolecular information (i.e., protein, deoxyribonucleic acid, and lipid content) contained in this spectral region for tissue characterization and diagnosis.[54, 64, 89] With the commonly used NIR 785 nm laser excitation source, however, intense tissue autofluorescence background and fused silica Raman signal arising from fiber-optic Raman probes also fall into the FP range. The tissue autofluorescence can severely interfere with the detection of the inherently weak FP Raman signals by saturating the CCD. This remains the primary concern in

certain organs such as lung, liver, and stomach.[72, 91] Recent reports have shown that HW (i.e., 2800 to 3700 cm^{-1}) Raman spectroscopy also contains valuable biomolecular information that can advantageously be used for diagnostic purposes.[114-117, 127, 153] The use of HW Raman spectroscopy is appealing due to the relatively intense tissue Raman signals generated (CH_2 and CH_3 moiety stretching vibrations of protein and lipids, OH stretching vibrations of water), as well as the considerably reduced fused silica fiber interferences and tissue autofluorescence that may allow a better assessment of the genuine tissue Raman signals.[114, 115] A recent study has established that the integrated FP and HW Raman technique offers complementary diagnostic information for increasing the accuracy and robustness of detecting precancer in cervical tissue.[116] To date, only very limited clinical work has been reported on the tissue Raman measurements covering both the FP and HW regions. The tissue Raman signals are measured either by successively switching the different laser excitation frequencies or by rotating the gratings for each spectral region,[153] which are not suitable for rapid *in vivo* measurements in clinical endoscopic settings. In this chapter, we report on the development of a novel multiplexed Raman spectroscopic technique that can simultaneously acquire both FP and HW *in vivo* tissue Raman spectra with high spectral resolution during endoscopy.

8.2 Raman instrumentation

Figure 8.1 shows the schematic of the multiplexed Raman spectroscopy system developed for simultaneous acquisition of the FP and HW Raman spectra in real-time under endoscopic image-guidance (i.e., WLR, NBI and AFI). The multiplexed Raman spectroscopy technique consists of a spectrum stabilized 785 nm diode laser (maximum output: 300 mW, B&W TEK Inc., Newark, Delaware), a transmissive

imaging spectrograph (Holospec f/1.8, Kaiser Optical Systems Inc., Ann Arbor, Michigan) equipped with a liquid nitrogen-cooled (-120°C), NIR-optimized, back-illuminated and deep depletion CCD camera (1340 × 400 pixels at 20 × 20 μm per pixel; Spec-10: 400BR/LN, Princeton Instruments, Roper Scientific Inc., Trenton, New Jersey). We integrate a VPH dual-transmission grating consisting of two hybrid VPH gratings (1400 and 1600 g/mm) (HoloPlex, Kaiser Optical Systems Inc., Ann Arbor, Michigan) into the spectrograph for Raman spectral dispersion.[154] The hybrid gratings are cemented closely together with a tilted angle of ~0.2- degree to achieve a separation between the low frequency and high frequency spectral components. The Bragg wavelengths of the hybrid gratings are tuned to two different spectral ranges (i.e., ~150 to 1950 cm^{-1} and 1750 to 3600 cm^{-1}) such that it disperses the tissue Raman spectra (i.e., FP and HW spectra) onto two separate vertical segments of the CCD, accordingly.

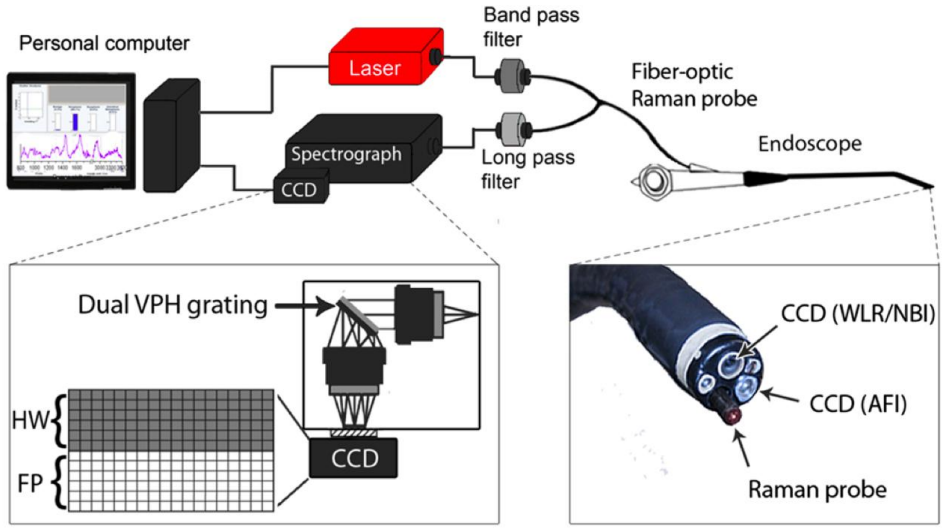


Figure 8.1 Schematic diagram of the rapid multiplexed Raman spectroscopy technique for simultaneous acquisition of both the fingerprint (FP) and high wavenumber (HW) Raman spectra under trimodal endoscopic imaging (i.e., white light reflectance (WLR), narrowband imaging (NBI), autofluorescence imaging (AFI)) guidance. A customized dual-transmission VPH grating is incorporated into the Raman system for dispersion of FP and HW Raman spectra onto different vertical segments of a CCD.

This multiplexed Raman spectroscopy technique based on a VPH dual-transmission grating permits simultaneous coverage of both the FP and HW spectral segments while maintaining a high spectral resolution of a single high-density grating. To correct for the image aberration in the transmissive spectrograph, a customized parabolic aligned fiber bundle (26 × 100 μm fibers, NA = 0.22) was coupled into the entrance slit of the spectrograph for significantly improving the signal-to-noise ratio (SNR) as well as the spectral resolution of the multiplexed Raman system as compared to a conventional straight slit imaging spectrograph.[54] This allows us to completely hardware bin the two separate CCD segments vertically for improving the SNR (of up to 14(=$200^{1/2}$) and spectral resolution (3 - 6 cm^{-1}) of the Raman spectra.[54] This is adequate for tissue spectroscopy that contains broad overlapping peaks. The two spectral segments are simultaneously readout so that the FP and HW regions can be spliced into a complete high-resolution broadband Raman spectrum covering ~150 to 3600 cm^{-1}. The rapid multiplexed Raman spectroscopy technique developed for endoscopy was wavelength calibrated to an accuracy (~2 cm^{-1}) using an argon/mercury spectral lamp (AR-1 and HG-1, Ocean Optics Inc., Dunedin, Florida) and the Raman spectrum of 4-acetamidophenol that exhibits strong well-defined peaks in the HW region at 2931 cm^{-1} and 3064 cm^{-1} (ASTM E1840 standard). To correct for the spectral response of the system, intensity calibration was performed using a standard tungsten lamp (RS-10, EG&G Gamma Scientific, San Diego, California) of the two distinct CCD segments separately. In this work, the 1.8 mm fiber-optic confocal Raman probe coupled with a 1.0 mm sapphire ball lens was used for epithelial tissue Raman measurements during endoscopic procedures. We have also developed customized software to process the two distinct CCD segments in real-time during clinical endoscopy, and the proper probe-tissue contact handling can be

verified on-line using outlier detection algorithms.

8.3 Data processing

The real-time data processing specially developed for this biomedical multiplexed FP and HW Raman spectroscopy technique includes silica fiber background subtraction, intensity and wavelength calibration, cosmic rays and signal saturation detection/rejection, autofluorescence background subtraction and linear Savitzky-Golay smoothing (5 pixel window).[43] Different tissue autofluorescence background subtraction schemes were employed for robust extraction of the tissue Raman signals. In the FP region (i.e., 800 to 1800 cm^{-1}), a 5^{th} order polynomial constrained to the lower portion of the FP Raman spectra is used and this polynomial is then subtracted from the measured spectrum to resolve the FP tissue Raman signal. In the HW region (i.e., 2800 to 3600 cm^{-1}), we found that 1^{st} order polynomials constrained to the lower portion of the HW Raman spectra are optimal for extracting the Raman spectra. All Raman spectra are also normalized to the integrated areas under the FP and HW Raman spectral regions, respectively, to reduce power density fluctuations associated with probe handling variations during clinical endoscopic examinations.[43]

8.4 Results and discussion

Figure 8.2 shows the full frame CCD image of a parabolic-configured fiber bundle illuminated with argon/mercury spectral lamps (AR-1 and HG-1, Ocean Optics Inc., Dunedin, Florida). With this specific fiber arrangement, the atomic emission lines are substantially straight, indicating effective image-aberration correction on the segmented CCD.[54] This in turn allows us to hardware bin the two well-defined CCD regions covering 1340 × 200 pixels without compromising spectral

resolutions or reducing the SNR ratio. The lower segment of the CCD array covers the FP region (i.e., ~150 to 1950 cm^{-1}), and we obtained a spectral resolution of ~6 cm^{-1} over the entire FP range using the 100 μm core diameter fibers. On the other hand, the upper segment of the CCD array covers the spectral range of 1750 to 3600 cm^{-1} comprising the nonspecific Raman region (~1800 to 2800 cm^{-1}) and the HW spectral region (i.e., 2800 to 3600 cm^{-1}) with a spectral resolution of ~3 cm^{-1}. Since the dispersion is holographically encoded in the grating for both FP and HW spectral regions, the signal magnitude and resolution of a single high-density grating are essentially maintained.

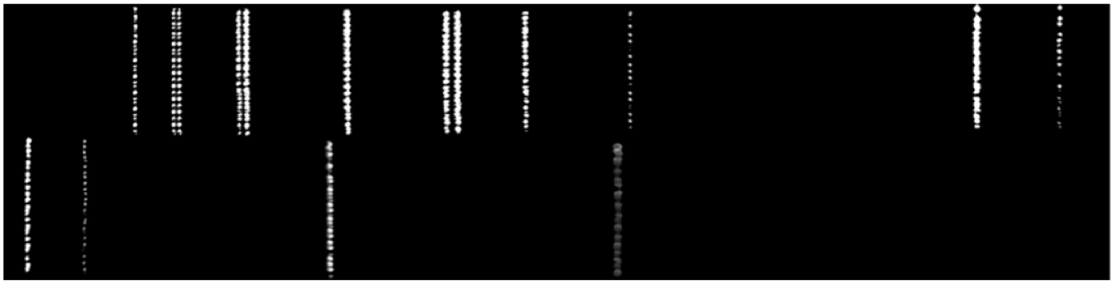


Figure 8.2 Image of a parabolic arranged fiber bundle (26 × 100 μm, NA = 0.22) illuminated with an argon/mercury lamp, illustrating the fully correction of image-aberration in spectrograph. The customized dual-transmission VPH grating efficiently disperses the FP and HW Raman spectra onto different vertical CCD segments. The lower segment covers the FP region ~150 to 1950 cm^{-1} (spectral resolution of 6 cm^{-1}) while the upper segment covers the nonspecific Raman range and the HW range from 1750 to 3600 cm^{-1} (spectral resolution of ~3 cm^{-1}).

We have illustrated the utility of the rapid multiplexed Raman spectroscopy system for real-time *in vivo* Raman measurements of epithelial tissue under wide-field endoscopic imaging (i.e., WLR/NBI/AFI) guidance. Figure 8.3 shows an example of *in vivo* Raman spectra acquired from different anatomical sites in the head and neck (i.e., attached gingiva, buccal mucosa, dorsal tongue, hard palate, and oropharynx) from a healthy volunteer under endoscopic imaging guidance. The FP and HW *in vivo* tissue Raman spectra can be acquired simultaneously with an integration time of 0.5 s and presented on the Raman endoscopy monitor in real-time. Highly resolved tissue

Raman peaks in the head and neck are observed in the FP range with tentative molecular assignments as follows: 853 cm^{-1} (υ(C-C) of proteins), 956 cm^{-1} (υ_s(P=O) of hydroxyapatite), 1004 cm^{-1} (υ_s(C-C) ring breathing of phenylalanine), 1078 cm^{-1} (υ(C-C)of lipids), 1265 cm^{-1} (amide III υ(C-N) and δ(N-H) of proteins), 1302 cm^{-1} (CH_3CH_2 twisting and wagging of proteins), 1445 cm^{-1} δ(CH_2) deformation of proteins and lipids), 1655 cm^{-1} (amide I υ(C=O) of proteins), and 1745 cm^{-1} υ(C=O) of lipids. Intense Raman peaks are also seen in the HW region such as 2850 and 2885 cm^{-1} (symmetric and asymmetric CH_2 stretching of lipids), 2940 cm^{-1} (CH_2 stretching of proteins) as well as the broad Raman band of water (OH stretching vibrations that peak at 3400 cm^{-1} in the 3100 to 3600 cm^{-1} region) (Table 1).

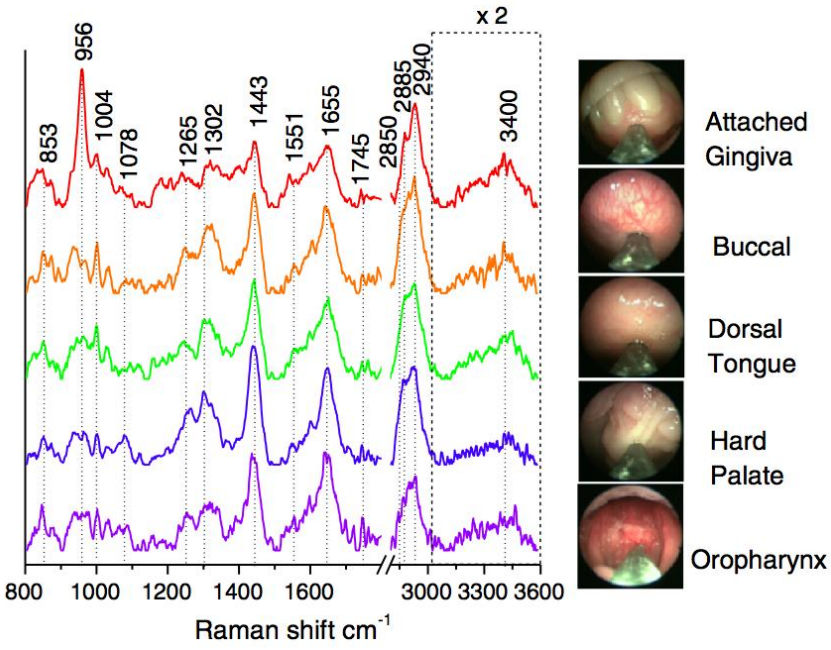


Figure 8.3 Example of *in vivo* Raman spectra acquired from different tissue sites (i.e., attached gingiva, buccal mucosa, dorsal tongue, hard palate and oropharynx) in the oral cavity from a healthy volunteer using the multiplexed Raman spectroscopy technique under endoscopic imaging guidance. The spectra are shifted vertically and normalized to the integrated areas in the FP and HW regions, respectively, for better comparisons of line shapes. All Raman spectra are acquired using an integration time of 0.5 s under the 785 nm illumination power of 1.5 W/cm^2. *In vivo* fiber-optic Raman endoscopic acquisitions under WLR imaging guidance are also shown.

The approximate SNR of 25 could be obtained from *in vivo* tissue using 1.0 s

integration time when the fiber-optic confocal Raman probe is in gentle contact with the buccal mucosa. Although the Raman band of water observed was relatively noisy due to the low quantum efficiency above 1090 nm of the CCD camera used, it may still contain important diagnostic information related to the local conformation and interactions of hydrogen-bonds in the cellular and extracellular space of tissue.[155] The OH stretching vibrations have been found to be associated with aquaporins and protein/water interactions in precancer and cancer tissues.[155] Therefore, compared with either FP or HW Raman spectroscopy technique alone, the multiplexed Raman spectroscopy technique utilizing a single excitation laser source together with a VPH dual-transmission grating provides both the FP and HW Raman spectra simultaneously (including OH stretching vibrations of water) with high spectral resolution (i.e., 3 to 6 cm^{-1}), paving the way for an improved tissue diagnosis and characterization *in vivo*.[116]

8.5 Conclusion

In summary, we report in this chapter on the development of a novel multiplexed Raman spectroscopy technique for simultaneous acquisition of the FP (800 to 1800 cm^{-1}) and HW (2800 to 3600 cm^{-1}) tissue Raman spectra *in vivo* with high spectral resolution (3 to 6 cm^{-1}). We demonstrate that high quality *in vivo* Raman spectra ranging from 800 to 3600 cm^{-1} of different tissue sites in the head and neck can be measured within 1.0 s integration time during endoscopic examination. The rapid multiplexed Raman spectroscopy technique with high spectral resolution developed in this work opens the opportunity for improving real-time *in vivo* tissue Raman diagnosis and characterization at endoscopy.

Chapter 9. Conclusions and future directions

We have herein advanced the field of biomedical Raman spectroscopy through a translational clinical study with the endpoint of applying fiber-optic Raman spectroscopy prospectively for *in vivo* endoscopic precancer and cancer diagnostics. In Chapter 4 we developed a novel fiber-optic endoscopic Raman spectroscopy technique using a customized fiber-optic Raman probe. This novel clinical instrument enabled tissue Raman spectra to be acquired within 0.5 sec which is necessary to meet the clinical requirements of rapid data acquisition. This is a factor of ~10 faster than any previously reported Raman techniques in the GI tract.[118] We demonstrated that this permits clinicians to nondestructively assess the endogenous biochemical and morphological information of GI tissues under the wide-field endoscopic imaging (WLR, NBI, and AFI) guidance. In Chapter 5 we elucidated the most prominent pathological features of intestinal-type gastric cancer and demonstrated the diagnostic capability of Raman spectroscopy for detecting precancers and cancers *in vivo*. To be able to apply fiber-optic Raman spectroscopy prospectively in clinical diagnostics it is necessary to apply sophisticated multivariate/chemometric analysis in real-time. Hence, we developed a novel on-line framework which enables real-time multivariate/chemometric analysis within 0.13 s. This is the first real-time Raman endoscopic technique that permits prospective and probabilistic tissue analysis based on a comprehensive patient database stored in the instrument itself. To improve the sensitivity for detecting epithelial lesions we also developed at depth-resolving fiber-optic Raman probe with confocal properties to selectively interrogate superficial tissues. This confocal-type Raman probe was tested in 77 esophagus patients, which represents the first prospective study reported to date in the literature and is therefore a milestone in clinical applications of Raman endoscopy. To further advance the field

we also developed a multiplexed FP and HW Raman endoscopy platform with improved resolution, broadband capability and real-time capability. This multiplexed Raman endoscope technique has potential to improve tissue characterization and diagnosis since complementary biomolecular information can be acquired from different spectral ranges in high resolution.

The use of fiber-optic Raman spectroscopy may include many clinical scenarios that require histopathologic biopsy samplings in screening of patients at higher risk of developing cancer. Ultimately, the targeting of dysplastic lesions and the discrimination between LGD and HGD, the clinical hallmarks in most GI carcinogenesis, will possibly be the key metric for defining the true clinical value of fiber-optic confocal Raman spectroscopy. Because dysplastic foci are prone to biopsy sampling errors, fiber-optic confocal Raman spectroscopy could be used to accurately pinpoint these high-risk tissue areas *in vivo* in real-time as well as examine those lesions that are indefinite for dysplasia or presenting with extensive inflammation where histopathologic manifestations are insufficient for clinical decision making. It may even be possible to detect field carcinogenesis or prognostic information in patients occurring before histopathologic manifestations. Fiber-optic confocal Raman spectroscopy could therefore be used to objectively stratify those patients who have higher risk of developing cancer and prompt increased frequency of surveillance or therapeutic intervention for the individual patients. These facets of Raman technology certainly warrant further investigations. Preemptive approaches or interventional procedures for HGD or intramucosal cancers are other scenarios where fiber-optic Raman spectroscopy could have a potential key role to play. Confocal Raman endoscopy offers the gastroenterologist a novel objective tool, which allows real-time

assessment of tissue pathology *in situ*, and therefore could be used to accurately define the resection margins of macroscopically inconspicuous dysplasia or cancers in patients undergoing surgical intervention (e.g., endoscopic mucosal resection, endoscopic submucosal dissection, ablation). This could enable complete excision and subsequent margin assessment, so that malignant progression can be efficiently prevented. Today, the necessity for histopathologic characterization of biopsy specimens hinders diagnosis and therapeutic intervention during the same endoscopic procedure. Confocal Raman technology opens up the possibility for combining final diagnosis and therapeutic eradication (e.g., endoscopic mucosal resection, endoscopic submucosal dissection) in a single procedure, thereby effectively reducing medical cost and burden associated with multiple endoscopic procedures. Fiber-optic confocal Raman spectroscopy also provides uninterrupted real-time diagnosis, which is straightforward to operate and requires no additional endoscopic training or administration of contrast agents. In our hospital, the confocal Raman technique is now routinely used by both experienced and novice clinicians to gather Raman spectral data without any difficulty. These clinical advantages will undoubtedly make Raman technique a competitive new modality with potential for rapid adoption and translation into gastroenterology. Because the basic principle of Raman spectroscopy is fundamentally different from endoscopic imaging modalities (e.g., WLR, NBI, AFI), Raman spectroscopy could possibly be complementary to conventional endoscopy. Fiber-optic Raman spectroscopy brings endoscopic examinations much closer to a molecular foundation and could challenge the current resource-intensive biopsy protocols. The endoscopist can now apply Raman endoscope to analyze the biomolecular signature of cells and tissues *in vivo*, thereby enabling noninvasive optical biopsies of suspicious lesions during ongoing endoscopy. The results of this

dissertation therefore illustrate the great potential of Raman endoscope technology for non-invasive, in vivo diagnosis and detection of pre-cancer and early cancer in the gastrointestinal system. We envisage that within the next five years, the fiber-optic Raman spectroscopy technique will be emerging as a new routine endoscopic tool for shaping *in vivo* biomedical diagnosis in hospitals.

9.1 Conclusions

Below we summarize the key achievements presented in this dissertation.

- A novel fiber-optic Raman spectroscopy technique was developed which enables acquisition of tissue Raman spectra in real-time from the upper GI tract. High-quality *in vivo* Raman spectra can be acquired from different sites of the upper GI (i.e., gastric and esophagus) in real-time during clinical endoscopic examinations. We applied the fiber-optic Raman technique for *in vivo* tissue characterization. The NNLSM coupled with a reference database of Raman active biochemicals (i.e., actin, histones, collagen, DNA, and triolein) was employed for semi-quantitative biomolecular modeling of tissue constituents in the upper GI. A total of 1189 *in vivo* Raman spectra were analyzed from different locations in the upper GI. The Raman spectra among the distal, middle, and proximal sites of the esophagus showed no significant interanatomical variability. The interanatomical variability of Raman spectra among normal gastric tissue (antrum, body, and cardia) was subtle compared to cancerous tissue transformation, whereas biomolecular modeling revealed significant differences between the two organs, particularly in the GE junction.

- To validate the potential of Raman endoscopy we assessed the capability for multi-class elucidation of intestinal-type gastric carcinogenesis sequence *in vivo* at endoscopy. We analyze 1277 *in vivo* Raman spectra from 83 gastric patients associated with intestinal-type carcinogenesis. The results show that *in vivo* Raman spectroscopy integrated with semi-quantitative spectral modeling (e.g., DNA, lipids, glycoprotein, proteins and blood) reveals the progressive changes of biochemical constituents in gastric tissue associated with preneoplastic and neoplastic transformation (i.e., IM, dysplasia and adenocarcinoma). Multi-class probabilistic partial least squares-discriminant analysis (PLS-DA) diagnostic algorithms based on *in vivo* Raman spectra are able to identify normal mucosa with sensitivity of 75.88% and specificity of 87.21%; IM with sensitivity of 46.67% and specificity of 87.55%; dysplasia with sensitivity of 83.33%; specificity of 95.80%, and adenocarcinoma with sensitivity of 84.91% and specificity 95.57%, respectively.

- Since data analysis of GI tissue Raman spectra has been limited to post-processing and off-line diagnostic algorithms we developed a novel on-line diagnostic framework that implements acquisition, preprocessing, outlier detection based on principal component analysis statistics (i.e., Hotelling's T^2 and Q-residuals) for tissue Raman spectra verification as well as probabilistic multivariate diagnostics. Free-running optical diagnosis and processing time of < 0.13 s can be achieved, which is a key achievement towards realizing real-time *in vivo* prospective tissue diagnostics during clinical endoscopic examination. We also demonstrated prospective cancer diagnosis in the gastric for the first time.

- To improve the sensitivity to precancer we developed a fiber-optic confocal-type

Raman technique. We demonstrated that real-time fiber-optic confocal Raman spectroscopy technique developed can be used prospectively in surveillance and screening of the patients with Barrett's esophagus *in vivo*. Confocal Raman spectra showed significant spectral differences ($p < 0.001$) among different esophageal tissue types (i.e., columnar-lined epithelium, nondysplastic BE and dysplastic BE). Fiber-optic confocal Raman spectroscopy provides a diagnostic sensitivity of 87.0% (67/77) and a specificity of 84.7% (610/720) for *in vivo* detection of dysplasia in 77 patients prospectively based on endogenous biomolecular signatures.

- To further advance the research, we developed a second generation multiplexed Raman spectroscopy technique for simultaneous acquisition of the FP (800 to 1800 cm^{-1}) and HW (2800 to 3600 cm^{-1}) tissue Raman spectra *in vivo* with high spectral resolution (3 to 6 cm^{-1}). We demonstrate that high quality *in vivo* Raman spectra ranging from 800 to 3600 cm^{-1} of different tissue sites in the head and neck can be measured within 1.0 s integration time during endoscopic examination. The rapid multiplexed Raman spectroscopy technique with high spectral resolution developed in this work opens the opportunity for improving real-time *in vivo* tissue Raman diagnosis and characterization at endoscopy.

9.2 Future directions

The work presented in the thesis mainly focused on the development and validation of fiber-optic Raman spectroscopy technique for endoscopic tissue diagnosis. There are methodological limitations in the current study as follows: (i) Biopsy was considered as gold standard. This implies that the co-targeting of the Raman measurement and biopsy sampling must be accurate in order to obtain an unbiased estimate of performance. Errors in the reported accuracies could be related to erroneous sampling which is an intrinsic uncertainty in this type of study. This effect could be quantitatively assessed through a multi-operator study using experienced and novice endoscopists. Secondly, although we used expert gastrointestinal pathologists it is well know that histopathological characterization of minute specimens can be highly subjective.[23] For instance, inter-observer disagreements of up to >50% have been reported for LGD in BE.[23] Hence, the diagnostic accuracy of Raman spectroscopy could be further improved using consensus pathology of several pathologists. (ii) Our dataset is limited in the number of patients with precancer (i.e., LGD and HGD) and cancer. This could give a misbalance in the accuracy reported. Also, we have not yet proved that Raman endoscopy is able to identify macroscopically invisible areas of dysplasia. (iii) Another limitation is that *in vivo* Raman spectra were acquired from a biased population of high-risk patients. (iv) Prospective diagnostics in our study was performed in a testing scenario. It is important to mention that the current Raman technology could aid endoscopists but might however not replace histopathology since Raman endoscopy provides little spatial information (e.g., invasion depth etc.). In order to assess the true clinical merit it is necessary to evaluate if Raman endoscopy can improve the guidance and yield of biopsies. Given the current stage of

this research, there are still issues to be addressed in order for the technology to gain clinical acceptance as a routine tool. This also includes standardization and conduction of multicenter randomized clinical trials. Some directions for further research are proposed as follows:

- Firstly, when using Raman spectroscopy to analyze tissue, the excitation laser power must be kept low to prevent thermal damage to tissue. Hence, light dosimetry for maximum permissible laser irradiation in internal organs is highly required, which is currently falling far behind the pace of Raman technology advancement. This is important with respect to health, safety and regulatory authorities. Standards need to be set for the exposure limits of NIR light to internal organs.

- Robust multivariate/chemometric approaches together with layered tissue modeling related to tissue optical properties (scattering and absorption coefficients, anisotropy) must also be developed for extracting biochemical and morphological diagnostic information in biomedical Raman spectroscopy for accurate tissue diagnosis and classification. Moreover, to better understand the spectral characteristics of *in vivo* Raman spectra and associated GI carcinogenesis, it would be informative to spectrally map histopathological tissue slides using confocal Raman microspectroscopy. Multivariate statistical analysis could then be employed i.e., PCA, biomolecular modelling, clustering and/or spectral unmixing methods etc. to extract biochemical and morphological information from GI mucosa. This could aid in better understanding of the distribution of Raman scatters in the gastric and esophagus for further improving diagnostic algorithm development and fiber-optic probe designs.

- Fiber-optic Raman spectroscopy has also mainly been limited to single system configurations and no attempts have been made to transfer into multi-centre clinical trials. Because Raman spectroscopy relies on extremely weak, well-resolved signals and the technique is very sensitive to instrumental changes such as interchanging of fiber-optic probes, this could limit the practical use in clinical trials and routine diagnostics. Instrument standardization therefore remains a very important subject for investigation in the future. Instrumental standardization also includes the assessment of probe durability, manufacturing and approval from regulatory authorities.

- One needs to validate the effect of confounding factors including probe pressure, temperature, ambient light conditions etc. It is however our practical experience that these are negligible compared to the tissue variance itself, but this needs to be scientifically proven. In addition the effects of age, gender, ethnicity must be evaluated for the technique to become a routine tool on hospitals in different parts of the world.

- Further, combining real-time fiber-optic Raman spectroscopy with other optical imaging techniques (e.g. hyper-spectral imaging, diffuse optical tomography (DOT), optical coherence tomography (OCT), confocal endomicroscopy imaging, and multiphoton imaging) and/or conventional biomedical imaging (e.g. ultra-sound, computed tomography (CT), and magnetic resonance imaging (MRI)) may also be the direction to explore for functional imaging at the molecular, cellular, tissue and even organ levels. The integration with OCT or confocal endomicroscopy could be particularly interesting directions to explore. OCT

provides micro-structural imaging with morphological information. The co-registration of Raman spectroscopy and OCT[156, 157] or confocal endomicroscopy could potentially enhance the tissue characterization and *in vivo* diagnosis. Future pragmatic developments will define the exact role of Raman endoscopy in medicine.

Bibliography

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: A Cancer Journal for Clinicians. 2011;61(2):69-90.

2. Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. Cancer Journal For Clinicians. 2010;60(5):277-300.

3. Leung WK, Wu MS, Kakugawa Y, Kim JJ, Yeoh KG, Goh KL, Wu KC, Wu DC, Sollano J, Kachintorn U, Gotoda T, Lin JT, You WC, Ng EK, Sung JJ. Screening for gastric cancer in Asia: current evidence and practice. Lancet Oncology. 2008;9(3):279-87. Epub 2008/03/01.

4. Brenner H, Rothenbacher D, Arndt V. Epidemiology of stomach cancer. Methods in molecular biology (Clifton, NJ). 2009;472:467-77. Epub 2008/12/25.

5. Jakszyn P, Gonzalez CA. Nitrosamine and related food intake and gastric and oesophageal cancer risk: a systematic review of the epidemiological evidence. World Journal of Gastroenterology. 2006;12(27):4296-303. Epub 2006/07/26.

6. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Research. 1992;52(24):6735-40. Epub 1992/12/15.

7. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N, Schlemper RJ. Helicobacter pylori infection and the development of gastric cancer. New England Journal of Medicine. 2001;345(11):784-9. Epub 2001/09/15.

8. Yeoh KG. How do we improve outcomes for gastric cancer? Journal of Gastroenteroly and Hepatology. 2007;22(7):970-2. Epub 2007/05/26.

9. Correa P, Piazuelo MB. Helicobacter pylori Infection and Gastric Adenocarcinoma. US gastroenterology & hepatology review. 2011;7(1):59-64. Epub

2011/08/23.

10. Tahara E. Genetic pathways of two types of gastric cancer. IARC Sci Publ. 2004(157):327-49. Epub 2004/04/02.

11. Graziano F, Humar B, Guilford P. The role of the E-cadherin gene (CDH1) in diffuse gastric cancer susceptibility: from the laboratory to clinical practice. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2003;14(12):1705-13. Epub 2003/11/25.

12. Zheng H, Takahashi H, Murai Y, Cui Z, Nomoto K, Miwa S, Tsuneyama K, Takano Y. Pathobiological characteristics of intestinal and diffuse-type gastric carcinoma in Japan: an immunostaining study on the tissue microarray. Journal of Clinical Pathology. 2007;60(3):273-7. Epub 2006/05/23.

13. Shaheen NJ, Richter JE. Barrett's oesophagus. The Lancet. 2009;373(9666):850-61.

14. Devesa SS, Blot WJ, Fraumeni JF. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. Cancer. 1998;83(10):2049-53.

15. Westhoff B, Brotze S, Weston A, McElhinney C, Cherian R, Mayo MS, Smith HJ, Sharma P. The frequency of Barrett's esophagus in high-risk patients with chronic GERD. Gastrointestinal endoscopy. 2005;61(2):226-31. Epub 2005/02/25.

16. Bersentes K, Fass R, Padda S, Johnson C, Sampliner RE. Prevalence of Barrett's esophagus in Hispanics is similar to Caucasians. Digestive Diseases and Sciences. 1998;43(5):1038-41. Epub 1998/05/20.

17. Souza RF, Krishnan K, Spechler SJ. Acid, bile, and CDX: the ABCs of making Barrett's metaplasia. American journal of physiology Gastrointestinal and liver physiology. 2008;295(2):G211-8. Epub 2008/06/17.

18. Wild CP, Hardie LJ. Reflux, Barrett's oesophagus and adenocarcinoma:

burning questions. Nature Reviews Cancer. 2003;3(9):676-84.

19. Overhiser AJ, Sharma P. Advances in endoscopic imaging: narrow band imaging. Reviews in gastroenterological disorders. 2008;8(3):186-93. Epub 2008/10/30.

20. Capelle L, Haringsma J, Vries A, Steyerberg E, Biermann K, Dekken H, Kuipers E. Narrow Band Imaging for the Detection of Gastric Intestinal Metaplasia and Dysplasia During Surveillance Endoscopy. Digestive Diseases and Sciences. 2010;55(12):3442-8.

21. Tada K, Oda I, Yokoi C, Taniguchi T, Sakamoto T, Suzuki H, Nonaka S, Yoshinaga S, Saito Y, Gotoda T. Pilot Study on Clinical Effectiveness of Autofluorescence Imaging for Early Gastric Cancer Diagnosis by Less Experienced Endoscopists. Diagnostic and Therapeutic Endoscopy. 2011;2011.

22. Reid BJ, Haggitt RC, Rubin CE, Roth G, Surawicz CM, Van Belle G, Lewin K, Weinstein WM, Antonioli DA, Goldman H, et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. Human pathology. 1988;19(2):166-78. Epub 1988/02/01.

23. Wang TD, Triadafilopoulos G, Crawford JM, Dixon LR, Bhandari T, Sahbaie P, Friedland S, Soetikno R, Contag CH. Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. Proceedings of the National Academy of Sciences. 2007;104(40):15864-9.

24. Kawamura A, Adachi K, Ishihara S, Katsube T, Takashima T, Yuki M, Amano K, Fukuda R, Yamashita Y, Kinoshita Y. Correlation between microsatellite instability and metachronous disease recurrence after endoscopic mucosal resection in patients with early stage gastric carcinoma. Cancer. 2001;91(2):339-45.

25. Huang Z, McWilliams A, Lui H, McLean DI, Lam S, Zeng H. Near-infrared

Raman spectroscopy for optical diagnosis of lung cancer. International Journal of Cancer. 2003;107(6):1047-52.

26. Teh SK, Zheng W, Ho KY, Teh M, Yeoh KG, Huang Z. Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue. British Journal of Cancer. 2008;98(2):457-65.

27. Teh SK, Zheng W, Ho KY, Teh M, Yeoh KG, Huang Z. Near-infrared Raman spectroscopy for gastric precancer diagnosis. Journal of Raman Spectroscopy. 2009;40(8):908-14.

28. Teh SK, Zheng W, Ho KY, Teh M, Yeoh KG, Huang Z. Near-infrared Raman spectroscopy for optical diagnosis in the stomach: Identification of Helicobacter-pylori infection and intestinal metaplasia. International Journal of Cancer. 2010;126(8):1920-7.

29. Teh SK, Zheng W, Ho KY, Teh M, Yeoh KG, Huang Z. Near-infrared Raman spectroscopy for early diagnosis and typing of adenocarcinoma in the stomach. British Journal of Surgery. 2010;97:550-7.

30. Teh SK, Zheng W, Ho KY, Teh M, Yeoh KG, Huang Z. Diagnosis of gastric cancer using near-infrared Raman spectroscopy and classification and regression tree techniques. Journal of Biomedical Optics. 2008;13(3):034013.

31. Teh SK, Zheng W, Lau DPC, Huang Z. Spectroscopic diagnosis of laryngeal carcinoma using near-infrared Raman spectroscopy and random recursive partitioning ensemble techniques. Analyst. 2009;134(6):1232-9.

32. So J, Rajnakova A, Chan YH, Tay A, Shah N, Salto-Tellez M, Teh M, Noriya U. Endoscopic Tri-Modal Imaging Improves Detection of Gastric Intestinal Metaplasia Among a High-Risk Patient Population in Singapore. Digestive Diseases and Sciences. 2013. Epub 2013/09/03.

33. Li M, Anastassiades CP, Joshi B, Komarck CM, Piraka C, Elmunzer BJ, Turgeon DK, Johnson TD, Appelman H, Beer DG, Wang TD. Affinity peptide for targeted detection of dysplasia in Barrett's esophagus. Gastroenterology. 2010;139(5):1472-80. Epub 2010/07/20.

34. Sharma P, Meining AR, Coron E, Lightdale CJ, Wolfsen HC, Bansal A, Bajbouj M, Galmiche JP, Abrams JA, Rastogi A, Gupta N, Michalek JE, Lauwers GY, Wallace MB. Real-time increased detection of neoplastic tissue in Barrett's esophagus with probe-based confocal laser endomicroscopy: final results of an international multicenter, prospective, randomized, controlled trial. Gastrointestinal endoscopy. 2011;74(3):465-72. Epub 2011/07/12.

35. Wallace MB, Crook JE, Saunders M, Lovat L, Coron E, Waxman I, Sharma P, Hwang JH, Banks M, DePreville M, Galmiche JP, Konda V, Diehl NN, Wolfsen HC. Multicenter, randomized, controlled trial of confocal laser endomicroscopy assessment of residual metaplasia after mucosal ablation or resection of GI neoplasia in Barrett's esophagus. Gastrointestinal endoscopy. 2012;76(3):539-47 e1. Epub 2012/07/04.

36. Li XD, Boppart SA, Van Dam J, Mashimo H, Mutinga M, Drexler W, Klein M, Pitris C, Krinsky ML, Brezinski ME, Fujimoto JG. Optical coherence tomography: advanced technology for the endoscopic imaging of Barrett's esophagus. Endoscopy. 2000;32(12):921-30. Epub 2001/01/09.

37. Almond LM, Hutchings J, Shepherd N, Barr H, Stone N, Kendall CA. Raman spectroscopy: a potential tool for early objective diagnosis of neoplasia in the oesophagus. Journal of Biophotonics. 2011;4(10):685-95. Epub 2011/08/10.

38. Evans JA, Poneros JM, Bouma BE, Bressner J, Halpern EF, Shishkov M, Lauwers GY, Mino-Kenudson M, Nishioka NS, Tearney GJ. Optical coherence

tomography to identify intramucosal carcinoma and high-grade dysplasia in Barrett's esophagus. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2006;4(1):38-43. Epub 2006/01/25.

39. Zhu Y, Terry NG, Wax A. Angle-resolved low-coherence interferometry: an optical biopsy technique for clinical detection of dysplasia in Barrett's esophagus. Expert review of gastroenterology & hepatology. 2012;6(1):37-41. Epub 2011/12/14.

40. Bensalah K, Peswani D, Tuncel A, Raman JD, Zeltser I, Liu H, Cadeddu J. Optical reflectance spectroscopy to differentiate benign from malignant renal tumors at surgery. Urology. 2009;73(1):178-81.

41. Chang SK, Mirabal YN, Atkinson EN, Cox DD, Malpica A, Follen M, Richards-Kortum RR. Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer. Journal of Biomedical Optics. 2005;10(2):024031.

42. Devpura S. Raman spectroscopy and diffuse reflectance spectroscopy for diagnosis of human cancer and acanthosis nigricans. 2012.

43. Duraipandian S, Bergholt MS, Zheng W, Ho KY, Teh M, Yeoh KG, Yan So JB, Huang Z. Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination. Journal of Biomedical Optics. 2012;17(8):081418.

44. Freeberg JA, Benedet JL, MacAulay C, West LA, Follen M. The performance of fluorescence and reflectance spectroscopy for the in vivo diagnosis of cervical neoplasia; point probe versus multispectral approaches. Gynecology and Oncology. 2007;107(1, Supplement 1):S248-S55.

45. Li QB, Xu Z, Zhang NW, Zhang L, Wang F, Yang LM, Wang JS, Zhou S,

Zhang YF, Zhou XS, Shi JS, Wu JG. In vivo and in situ detection of colorectal cancer using Fourier transform infrared spectroscopy. World Journal of Gastroenterology. 2005;11(3):327-30. Epub 2005/01/08.

46. Lim L, Nichols B, Rajaram N, Tunnell JW. Probe pressure effects on human skin diffuse reflectance and fluorescence spectroscopy measurements. Journal of Biomedical Optics. 2011;16(1):011012.

47. Lovat LB, Johnson K, Mackenzie GD, Clark BR, Novelli MR, Davies S, O'Donovan M, Selvasekar C, Thorpe SM, Pickard D, Fitzgerald R, Fearn T, Bigio I, Bown SG. Elastic scattering spectroscopy accurately detects high grade dysplasia and cancer in Barrett's oesophagus. Gut. 2006;55(8):1078-83. Epub 2006/02/14.

48. Qiu L, Pleskow DK, Chuttani R, Vitkin E, Leyden J, Ozden N, Itani S, Guo L, Sacks A, Goldsmith JD, Modell MD, Hanlon EB, Itzkan I, Perelman LT. Multispectral scanning during endoscopy guides biopsy of dysplasia in Barrett's esophagus. Nature Medicine. 2010;16(5):603-6.

49. Mayinger B, Horner P, Jordan M, Gerlach C, Horbach T, Hohenberger W, Hahn E. Endoscopic fluorescence spectroscopy in the upper GI tract for the detection of GI cancer: initial experience. The American journal of gastroenterology. 2001;96(9):2616-21.

50. Lin B, Bergholt MS, Lau DP, Huang Z. Diagnosis of early stage nasopharyngeal carcinoma using ultraviolet autofluorescence excitation-emission matrix spectroscopy and parallel factor analysis. Analyst. 2011;136(19):3896-903.

51. Shao X, Zheng W, Huang Z. Near-infrared autofluorescence spectroscopy for in vivo identification of hyperplastic and adenomatous polyps in the colon. Biosensors & Bioelectronics. 2011;30(1):118-22.

52. Shao X, Zheng W, Huang Z. In vivo diagnosis of colonic precancer and cancer

using near-infrared autofluorescence spectroscopy and biochemical modeling. Journal of Biomedical Optics. 2011;16(6):067005.

53. Shao X, Zheng W, Huang Z. Polarized near-infrared autofluorescence imaging combined with near-infrared diffuse reflectance imaging for improving colonic cancer detection. Optics Express. 2010;18(23):24293-300.

54. Huang Z, Zeng H, Hamzavi I, McLean DI, Lui H. Rapid near-infrared Raman spectroscopy system for real-time in vivo skin measurements. Optics Letters. 2001;26(22):1782-4.

55. Zavaleta CL, Garai E, Liu JTC, Sensarn S, Mandella MJ, Van de Sompel D, Friedland S, Van Dam J, Contag CH, Gambhir SS. A Raman-based endoscopic strategy for multiplexed molecular imaging. PNAS. 2013.

56. Raman CV, Krishnan KS. A new type of secondary radiation. Nature. 1928;121(3048):501-2.

57. McCreery RL. Chemical Analysis. Winefordner JD, editor. New York, USA: Wiley-Interscience; 2000. 420 p.

58. Long DA. Vibrational Raman Scattering. The Raman Effect: John Wiley & Sons, Ltd; 2002. p. 85-152.

59. Hassing S. Molecules and radiation. Lecture notes2009.

60. Lam T. A new era in affordable Raman spectroscopy. Raman Technology for Today's Spectroscopists. 2004:30-7.

61. Bandyopadhyay AK, Dilawar N, Vijayakumar A, Varandani D, Singh D. A low cost laser-raman spectrometer. Bulletin of Materials Science. 1998;21(5):433-8.

62. Zeng H, Zhao J, Short M, McLean DI, Lam S, McWilliams A, Harvey L. Raman spectroscopy for in vivo tissue analysis and diagnosis, from instrument development to clinical applications. Journal of Innovative Optical Health Sciences.

2008;1(1):95-106.

63. Utzinger U, Richards-Kortum RR. Fiber optic probes for biomedical optical spectroscopy. Journal of Biomedical Optics. 2003;8(1):121-47.

64. Huang Z, Teh SK, Zheng W, Mo J, Lin K, Shao X, Ho KY, Teh M, Yeoh KG. Integrated Raman spectroscopy and trimodal wide-field imaging techniques for real-time in vivo tissue Raman measurements at endoscopy. Optics Letters. 2009;34(6):758-60.

65. Latka I, Dochow S, Krafft C, Dietzek B, Popp J. Fiber optic probes for linear and nonlinear Raman applications – Current trends and future development. Laser & Photonics Reviews. 2013.

66. Motz JT, Hunter M, Galindo LH, Gardecki JA, Kramer JR, Dasari RR, Feld MS. Optical fiber probe for biomedical Raman spectroscopy. Applied Optics. 2004;43(3):542-54.

67. Zhang G, Demos SG, Alfano RR. Raman spectra of biomedical samples using optical fiber probes. Proc SPIE. 1997;2976:2-9.

68. Day JCC, Bennett R, Smith B, Kendall CA, Hutchings J, Meaden GM, Born C, Yu S, Stone N. A miniature confocal Raman probe for endoscopic use. Physics in medicine and biology. 2009;54(23):7077-87.

69. Shim MG, Wilson BC, Marple E, Wach M. Study of fiber-optic probes for in vivo medical Raman spectroscopy. Applied Spectroscopy. 1999;53(6):619-27.

70. Shim MG, Song LMWK, Marcon NE, Wilson BC. *In vivo* near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy. Photochemistry and Photobiology. 2000;72(1):146-50.

71. Song LMWK, Molckovsky A, Wang KK, Burgart LJ, Dolenko B, Somorjai RL, Wilson BC. Diagnostic potential of Raman spectroscopy in Barrett's esophagus.

Proc SPIE. 2005;5692:140-6.

72. Short MA, Lam S, McWilliams A, Zhao J, Lui H, Zeng H. Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers. Optics Letters. 2008;33(7):711-3.

73. Krafft C, Dochow S, Latka I, Becker M, Spittel R, Kobelke J, Schuster K, Rothardt M, Popp J. Multi-core fiber with integrated fiber Bragg grating for background free Raman sensing. Optics Express. 2013;20(18):20156-69.

74. American National Standard for the Safe Use of Lasers. Washington, D.C.: American National Standards Institute, 1986 ANSI Standard 2136.1-1986.

75. Miller CE. Chemometrics in Process Analytical Technology (PAT). In: Bakeev KA, editor. Process Analytical Technology. second edition ed. New York, USA: John Wiley & Sons Ltd; 2010. p. 353-438.

76. Ehrentreich F, Sümmchen L. Spike Removal and Denoising of Raman Spectra by Wavelet Transform Methods. Analytical Chemistry. 2001;73(17):4364-73.

77. Zhao J, Lui H, McLean DI, Zeng H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. Applied spectroscopy. 2007;61(11):1225-32.

78. Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Yan So JB, Huang Z. Combining near-infrared-excited autofluorescence and Raman spectroscopy improves in vivo diagnosis of gastric cancer. Biosensors & Bioelectronics. 2011;26(10):4104-10.

79. Gang L, editor. Removing Background of Raman Spectrum Based on Wavelet Transform. Future Computer and Communication, 2009 FCC '09 International Conference on; 2009 6-7 June 2009.

80. Barman I, Kong CR, Singh GP, Dasari RR. Effect of photobleaching on

calibration model development in biological Raman spectroscopy. Journal of Biomedical Optics. 2011;16(1):011004. Epub 2011/02/02.

81. Mo J, Zheng W, Low JJH, Ng J, Ilancheran A, Huang Z. Near-infrared autofluorescence spectroscopy for in vivo diagnosis of cervical precancer. Proc SPIE. 2008;6853:685317. Epub 1.

82. Brereton R. Chemometrics for Pattern Recognition2009.

83. Widjaja E, Zheng W, Huang Z. Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines. International Journal of Oncology. 2008;32(3):653-62.

84. Duraipandian S, Zheng W, Ng J, Low JJH, Ilancheran A, Huang Z. In vivo diagnosis of cervical precancer using Raman spectroscopy and genetic algorithm techniques. Analyst. 2011;136(20):4328-36.

85. Leardi R. Application of genetic algorithm-PLS for feature selection in spectral data sets. Journal of Chemometrics. 2000;14(5-6):643-55.

86. Leardi R. Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection. Journal of Chemometrics. 1994;8(1):65-79.

87. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. Journal of Chemometrics. 1992;6(5):267-81.

88. Leardi R, Lupianez Gonzalez A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemometrics and Intelligent Laboratory Systems. 1998;41(2):195-207.

89. Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Yan So JB, Huang Z. In vivo diagnosis of gastric cancer using raman endoscopy and ant colony optimization techniques. International journal of cancer. 2011;128(11):2673-80.

90. Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Yan So JB, Huang Z. Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach. Analyst. 2010;135(12):3162-8.

91. Bergholt MS, Zheng W, Ho KY, Teh M, Yeoh KG, So JBY, Shabbir A, Huang Z. Fiber-optic Raman spectroscopy probes gastric carcinogenesis in vivo at endoscopy. Journal of Biophotonics. 2013;6(1):49-59.

92. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36. Epub 1982/04/01.

93. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari RR, Crowe JP, Feld MS. In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy. Cancer Research. 2006;66(6):3317-22.

94. Shetty G, Kendall CA, Shepherd N, Stone N, Barr H. Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. British Journal of Cancer. 2006;94(10):1460-4.

95. Almond LM, Hutchings J, Barr H, Day JCC, Stone N, Kendall CA. Development of fibre-optic Raman probes for in vivo diagnosis of upper gastrointestinal cancers. Spectroscopy Europe. 2011;23(3):6-12.

96. Stone N, Kendall CA, Shepherd N, Crow P, Barr H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. Journal of Raman Spectroscopy. 2002;33(7):564-73.

97. Stone N, Kendall CA, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. Faraday Discussions. 2004;126:141-57.

98. Huang Z, Zheng W, Colin S. Near-Infrared Raman Spectroscopy for Colonic

Cancer Diagnosis. Proc SPIE. 2005;5862.

99. Shim MG, Song LMWK, Marcon NE, Hassaram S, Wilson BC. Assessment of ex-vivo and in-vivo near-infrared Raman spectroscopy for the classification of dysplasia within Barrett's esophagus. Proc SPIE. 2000;3918:114-9.

100. Shim MG, Wilson BC. Development of an in vivo Raman spectroscopic system for diagnostic applications. Journal of Raman Spectroscopy. 1997;28(2-3):131-42.

101. Shim MG, Wilson BC. The Effects of ex vivo Handling Procedures on the Near-Infrared Raman Spectra of Normal Mammalian Tissues. Photochemistry and Photobiology. 1996;63(5):662-71.

102. Shim MG, Wilson BC, Marple E, Wach ML. Evaluation of fiber optic probes for in-vivo Raman spectroscopy. 1998:208-17.

103. Krafft C, Neudert L, Simat T, Salzer R. Near infrared Raman spectra of human brain lipids. Spectrochimica Acta Part A: Molecular Spectroscopy. 2005;61(7):1529-35. Epub 2005/04/12.

104. Towler MR, Wren A, Rushe N, Saunders J, Cummins NM, Jakeman PM. Raman spectroscopy of the human nail: a potential tool for evaluating bone health? Journal of Materials Science. 2007;18(5):759-63. Epub 2006/12/01.

105. Binoy J, Abraham JP, Joe IH, Jayakumar VS, Pettit GR, Nielsen OF. NIR-FT Raman and FT-IR spectral studies and ab initio calculations of the anti-cancer drug combretastatin-A4. Journal of Raman Spectroscopy. 2004;35(11):939-46.

106. Bergholt MS, Zheng W, Huang Z. Characterizing variability in in vivo Raman spectroscopic properties of different anatomical sites of normal tissue in the oral cavity. Journal of Raman Spectroscopy. 2012;43(2):255-62.

107. Morris MD, Mandair GS. Raman assessment of bone quality. Clinical

orthopaedics and related research. 2011;469(8):2160-9. Epub 2010/12/01.

108. Shao J, Lin M, Li Y, Li X, Liu J, Liang J, Yao H. In Vivo Blood Glucose Quantification Using Raman Spectroscopy. PLoS ONE. 2012;7(10):e48127.

109. Chan JW, Taylor DS, Zwerdling T, Lane SM, Ihara K, Huser T. Micro-Raman Spectroscopy Detects Individual Neoplastic and Normal Hematopoietic Cells. Biophysical Journal. 2006;90(2):648-56.

110. Notingher I, Green C, Dyer C, Perkins E, Hopkins N, Lindsay C, Hench LL. Discrimination between ricin and sulphur mustard toxicity in vitro using Raman spectroscopy. Journal of the Royal Society, Interface / the Royal Society 2004;1(1):79-90. Epub 2006/07/20.

111. Cheng WT, Liu MT, Liu HN, Lin SY. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. Microscopy research and technique. 2005;68(2):75-9. Epub 2005/10/18.

112. Bergholt MS, Lin K, Zheng W, Lau DPC, Huang Z. In vivo, real-time, transnasal, image-guided Raman endoscopy: defining spectral properties in the nasopharynx and larynx. Journal of Biomedical Optics. 2012;17(7):077002.

113. Garcia-Flores A, Raniero L, Canevari R, Jalkanen K, Bitar R, Martinho H, Martin A. High-wavenumber FT-Raman spectroscopy for in vivo and ex vivo measurements of breast cancer. Theoretical Chemistry Accounts. 2011;130(4):1231-8.

114. Koljenovic S, Schut TCB, Wolthuis R, de Jong BWD, Santos L, Caspers PJ, Kros JM, Puppels GJ. Tissue characterization using high wave number Raman spectroscopy. Journal of Biomedical Optics. 2005;10(3):031116.

115. Lin K, Lau DPC, Huang Z. Optical diagnosis of laryngeal cancer using high wavenumber Raman spectroscopy. Biosensors & Bioelectronics. 2012;35(1):213-7.

116. Duraipandian S, Zheng W, Ng J, Low JJH, Ilancheran A, Huang Z.

Simultaneous fingerprint and high-wavenumber confocal Raman spectroscopy enhances early detection of cervical precancer in vivo. Analytical Chemistry. 2012;84(14):5913-9.

117. Mo J, Zheng W, Low JJH, Ng J, Ilancheran A, Huang Z. High wavenumber Raman spectroscopy for in vivo detection of cervical dysplasia. Analytical Chemistry. 2009;81(21):8908-15.

118. Molckovsky A, Song LMWK, Shim MG, Marcon NE, Wilson BC. Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. Gastrointestinal endoscopy. 2003;57(3):396-402.

119. Bergholt MS, Zheng W, Huang Z. Development of a multiplexing fingerprint and high wavenumber Raman spectroscopy technique for real-time in vivo tissue Raman measurements at endoscopy. Journal of Biomedical Optics. 2013;18(3):030502. Epub 2013/03/02.

120. Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Huang Z. In vivo Raman spectroscopy integrated with multimodal endoscopic imaging for early diagnosis of gastric dysplasia. Proc SPIE. 2010;7560(1):756003.

121. Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Yan So JB, Huang Z. Characterizing variability in in vivo Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection. Journal of Biomedical Optics. 2011;16(3):037003.

122. Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Yan So JB, Huang Z. In vivo diagnosis of esophageal cancer using image-guided Raman endoscopy and biomolecular modeling. Technology in cancer research & treatment. 2011;10(2):103-12.

123. Huang Z, Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG. In vivo early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy. Journal of Biomedical Optics. 2010;15(3):037017.

124. Huang Z, Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, Champion P, Ziegler L, editors. Image-Guided Raman Spectroscopy For In Vivo Diagnosis of Gastric Precancer At Gastroscopy. 2010.

125. Wang J, Bergholt MS, Zheng W, Huang Z. Development of a beveled fiber-optic confocal Raman probe for enhancing in vivo epithelial tissue Raman measurements at endoscopy. Opt Lett. 2013;accepted.

126. Bergholt MS, Zheng W, Ho KY, Teh M, Yeoh KG, Yan So JB, Shabbir A, Huang Z. Fiberoptic Confocal Raman Spectroscopy for Real-Time In Vivo Diagnosis of Dysplasia in Barrett's Esophagus. Gastroenterology. 2014;146(1):27-32. Epub 2013/11/13.

127. Short MA, Lam S, McWilliams AM, Ionescu DN, Zeng H. Using laser Raman spectroscopy to reduce false positives of autofluorescence bronchoscopies: a pilot study. Journal of Thoracic Oncology. 2011;6(7):1206-14. Epub 2011/08/19.

128. Draga ROP, Grimbergen MCM, Vijverberg PLM, van Swol CFP, Jonges TGN, Kummer JA, Ruud Bosch JLH. In vivo bladder cancer diagnosis by high-volume Raman spectroscopy. Analytical Chemistry. 2010;82(14):5993-9.

129. Zeng H, Petek M, Zorman MT, McWilliams A, Palcic B, Lam S. Integrated endoscopy system for simultaneous imaging and spectroscopyfor early lung cancer detection. Optics Letters. 2004;29(6):587-9.

130. Owen DA. Normal histology of the stomach. The American journal of surgical pathology. 1986;10(1):48-61. Epub 1986/01/01.

131. Takubo. Structure of the Esophagus. Pathology of the Esophagus: Springer

Japan; 2009. p. 8-45.

132. Huang Z, Teh SK, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG. In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy. Biosensors & Bioelectronics. 2010;26(2):383-9.

133. F. P. Incropera D. P. D. Witt. Fundamentals of Heat and Mass Transfer: , John Wiley and Sons, New York 1990.

134. Torres JH, Motamedi M, Pearce JA, Welch AJ. Experimental evaluation of mathematical models for predicting the thermal response of tissue to laser irradiation. Applied Optics. 1993;32(4):597-606.

135. Diem M, Boydston-White S, Chiriboga L. Infrared Spectroscopy of Cells and Tissues: Shining Light onto a novel Subject. Applied Spectroscopy. 1999;53(4):148A-61A.

136. Hopwood D, Logan KR, Coghill G, Bouchier IA. Histochemical studies of mucosubstances and lipids in normal human oesophageal epithelium. The Histochemical journal. 1977;9(2):153-61. Epub 1977/03/01.

137. Short KW, Carpenter S, Freyer JP, Mourant JR. Raman spectroscopy detects biochemical changes due to proliferation in mammalian cell cultures. Biophysical Journal. 2005;88(6):4274-88.

138. Leung WK, Sung JJ. Review article: intestinal metaplasia and gastric carcinogenesis. Aliment Pharmacol Ther. 2002;16(7):1209-16. Epub 2002/07/30.

139. Falk GW. Autofluorescence endoscopy. Gastrointestinal endoscopy clinics of North America. 2009;19(2):209-20. Epub 2009/05/09.

140. Liu Q, Teh M, Ito K, Shah N, Ito Y, Yeoh KG. CDX2 expression is progressively decreased in human gastric intestinal metaplasia, dysplasia and cancer. Modern pathology : an official journal of the United States and Canadian Academy of

Pathology, Inc. 2007;20(12):1286-97. Epub 2007/10/02.

141. Mao XY, Wang XG, Lv XJ, Xu L, Han CB. COX-2 expression in gastric cancer and its relationship with angiogenesis using tissue microarray. World Journal of Gastroenterology. 2007;13(25):3466-71. Epub 2007/07/31.

142. Joypaul BV, Newman EL, Hopwood D, Grant A, Qureshi S, Lane DP, Cuschieri A. Expression of p53 protein in normal, dysplastic, and malignant gastric mucosa: an immunohistochemical study. The Journal of pathology. 1993;170(3):279-83. Epub 1993/07/01.

143. Zhao J, Lui H, Mclean DI, Zeng H. Real-time Raman spectroscopy for noninvasive in vivo skin analysis and diagnosis. In: Campolo D, editor. New Developments In Biomedical Engineering. Vukovar, Croatia: InTech; 2010. p. 455-74.

144. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe JP, Dasari RR, Feld MS. Diagnosing breast cancer by using Raman spectroscopy. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(35):12371-6.

145. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe JP, Dasari RR, Feld MS. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. Cancer Research. 2002;62(18):5375-80.

146. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Shenk R, Wang N, Dasari RR, Fitzmaurice M, Feld MS. Diagnosing breast cancer using Raman spectroscopy: prospective analysis. Journal of Biomedical Optics. 2009;14(5):054023.

147. Motz JT, Fitzmaurice M, Miller A, Gandhi SJ, Haka AS, Galindo LH, Dasari RR, Kramer JR, Feld MS. In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque. Journal of Biomedical Optics. 2006;11(2):021003-.

148. Motz JT, Gandhi SJ, Scepanovic OR, Haka AS, Kramer JR, Dasari RR, Feld

MS. Real-time Raman system for in vivo disease diagnosis. Journal of Biomedical Optics. 2005;10(3):031113.

149. Cogdill RP, Anderson CA, Drennen III JK. Process analytical technology case study, part III: calibration monitoring and transfer. AAPS PharmSciTech. 2005;6(2):E284-97. Epub 2005/12/16.

150. Ortiz-Estarelles O, Martin-Biosca Y, Medina-Hernandez MJ, Sagrado S, Bonet-Domingo E. Multivariate data analysis of quality parameters in drinking water. Analyst. 2001;126(1):91-6.

151. Martin EB, Morris AJ, Zhang J. Process performance monitoring using multivariate statistical process control. IEE Proceedings-Control Theory and Applications 1996;143(2):132-44.

152. Wang J, Bergholt MS, Zheng W, Huang Z. Development of a beveled fiber-optic confocal Raman probe for enhancing in vivo epithelial tissue Raman measurements at endoscopy. Optics Letters. 2013;38(13):2321-3.

153. Chau AH, Motz JT, Gardecki JA, Waxman S, Bouma BE, Tearney GJ. Fingerprint and high-wavenumber Raman spectroscopy in a human-swine coronary xenograft in vivo. Journal of Biomedical Optics. 2008;13(4):040501. Epub 2008/11/22.

154. Owen H. The Impact of Volume Phase Holographic Filters and Gratings on the Development of Raman Instrumentation. Journal of Chemical Education. 2007;84(1):61.

155. Chagas e Silva de Carvalho L, Sato É, Almeida J, da Silva Martinho H. Diagnosis of inflammatory lesions by high-wavenumber FT-Raman spectroscopy. Theoretical Chemistry Accounts. 2011;130(4-6):1221-9.

156. Patil CA, Bosschaart N, Keller MD, van Leeuwen TG, Mahadevan-Jansen A.

Combined Raman spectroscopy and optical coherence tomography device for tissue characterization. Optics Letters. 2008;33(10):1135-7.

157. Khan KM, Krishna H, Majumder SK, Rao KD, Gupta PK. Depth-sensitive Raman spectroscopy combined with optical coherence tomography for layered tissue analysis. Journal of Biophotonics. 2013:1-9.