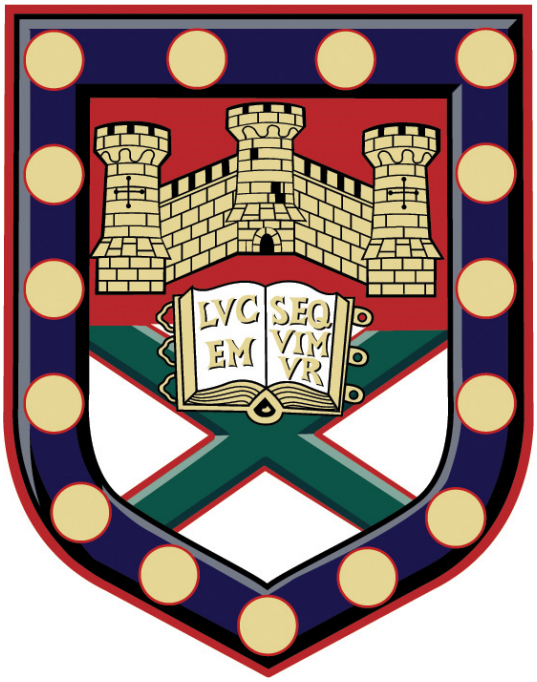
UNIVERSITY OF EXETER



DETECTION OF BARRETT'S NEOPLASIA WITH VIBRATIONAL SPECTROSCOPY

OLIVER J OLD

MD THESIS

SUPERVISORS: PROF. AC SHORE, PROF. N STONE, DR C KENDALL, MR LM ALMOND AND PROF. H BARR.

NOVEMBER 2015

DETECTION OF BARRETT'S NEOPLASIA WITH VIBRATIONAL SPECTROSCOPY

Submitted by Oliver Jamieson Old to the University of Exeter

As a thesis for the degree of Doctor of Medicine

November 2015

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……Oliver Old……..

ABSTRACT

Early detection of Barrett's oesophagus and associated neoplasia is key to preventing progression to oesophageal adenocarcinoma. Improving surveillance and introducing population screening for Barrett's are major goals of current research: this project aimed to apply emerging techniques in vibrational spectroscopy to these problems.

Fourier transform infrared (FTIR) mapping was used to develop an automated histology tool for detection of Barrett's and Barrett's neoplasia in tissue biopsies. 45 FTIR maps were measured from 22 tissue samples from 19 patients. Principal component analysis (PCA) fed linear discriminant analysis (LDA) was used to build classification models based on spectral differences, tested using leave one sample out cross validation (LOSOCV). Classification of normal squamous samples versus 'abnormal' samples (any stage of Barrett's) was performed with 100% sensitivity and specificity. Using a 3-group model to differentiate normal squamous, non-dysplastic Barrett's and neoplastic Barrett's (dysplasia or adenocarcinoma), neoplastic Barrett's was identified with 95.6% sensitivity and 86.4% specificity.

Non-endoscopic cell collection devices have recently been developed for population screening for Barrett's oesophagus. A further aim of this project was to evaluate FTIR for classification of oesophageal cells. Cytology brushings were collected at endoscopy, cytospun onto slides and FTIR maps measured. Cytology review and contemporaneous histology was used to inform a training dataset containing 141 cells from 17 patients. A classification model was constructed using PCA-fed LDA. Applying this training model to the entire dataset of 115 FTIR maps from 66 patients, whole samples were classified with

sensitivity and specificity respectively as follows: normal squamous 79.0% and 77.0%, non-dysplastic Barrett's 31.3% and 100%, and neoplastic Barrett's 83.3% and 54.2%.

Raman spectroscopy was also evaluated as a tool for tissue diagnosis, but several strands of enquiry were limited by instrument problems.

FTIR mapping could be used as an accurate, automated tool for processing biopsies in Barrett's surveillance. Analysis of oesophageal cell samples can be performed using FTIR with reasonable sensitivity for Barrett's neoplasia, though poor specificity with the current technique.

ACKNOWLEDGEMENTS

I would like to thank the Royal College of Surgeons of England, who awarded me a Research Fellowship that provided funding for my first year of study, and a number of further opportunities within the College that arose as a result of this Fellowship.

Secondly, I must thank everyone at the Biophotonics Research Unit in Gloucester. The Biophotonics Research Unit funded my second year of study, and every member of the team has been a source of help and support during the project. Special thanks are due to Martin Isabelle and Gavin Lloyd for their efforts and advice with data analysis, and particularly Gavin's contribution to developing the binary mask for cellular analysis.

Thanks also to the Physics Department at the University of Exeter, and in particular Krupakar Nallala and Kelly Curtis, for help with experimental work performed there.

My supervisors have been a pleasure to work with, and I must thank each for their own contribution to the project: my predecessor Max Almond for his work leading to this project, Catherine Kendall for her dedication and attention to detail, Nick Stone for his endless ideas and overall vision, Angela Shore for adding much needed external perspective, and finally Professor Barr for his tireless enthusiasm, which has been an inspiration.

Finally, I would like to thank my wife Phoebe, who has been a constant source of strength and support, as always.

This thesis is dedicated to Rory.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AC	Adenocarcinoma
ACG	American College of Gastroenterologists
AFI	Autofluorescence imaging
AGA	American Gastroenterological Society
APC	Argon plasma coagulation
AspECT	Aspirin Esomeprazole Chemoprevention Trial
BE/BO	Barrett's oesophagus
BRIDE	Barrett's Radiofrequency Intervention for Dysplasia by Endoscopy
BSG	British Society of Gastroenterology
CARS	Coherent anti-Stokes Raman spectroscopy
ChoPIN	ChemoPrevention of Intestinal Neoplasia
CLE	Confocal laser endomicroscopy
CLO	Columnar-lined oesophagus
CPI	Cross-polarized imaging
DYS	Dysplasia
EAGLE	Esophageal Adenocarcinoma Genetics Linkage
EMSC	Extended multiplicative scatter correction
EPSS	Endoscopic polarised scanning spectroscopy

ESS	Elastic scattering spectroscopy
FICE	Fuji Intelligent Chromo Endoscopy
FTIR	Fourier Transform Infrared
GI	Gastrointestinal
GORD	Gastro-oesophageal reflux disease
H2RA	Histamine-2 receptor antagonists
HGD	High grade dysplasia
HRE	High resolution endoscope
HRME	High resolution microendoscopy
IM	Intestinal metaplasia
LGD	Low grade dysplasia
MMP	Matrix metalloproteinase
NBI	Narrow band imaging
NICE	National Institute for Health and Clinical Excellence
NSAIDs	Non-steroidal anti-inflammatory drugs
NSQ	Normal squamous
OAC	Oesophageal adenocarcinoma
OCT	Optical coherence tomography
OGD	Oesophagogastroduodenoscopy

RCT	Randomised controlled trials
SCP	Spectral cytopathology
SD	Standard deviation
SERS	Surface enhanced Raman spectroscopy
SORS	Spatially offset Raman spectroscopy
SRS	Stimulated Raman spectroscopy
SSBO	Short segment Barrett’s oesophagus
SNP	Single nucleotide polymorphism
TERS	Tip-enhanced Raman spectroscopy
TNM	Tumour-node-metastasis
VFI	Vital dye fluorescence imaging
WLE	White light endoscopy

LIST OF APPENDICES

RESEARCH CONTEXT AND AIMS

Early diagnosis is the key to improving outcomes from oesophageal adenocarcinoma: although Barrett's oesophagus is a recognised pre-malignant condition with established surveillance programmes for identified patients, over 90% of oesophageal adenocarcinomas present as symptomatic tumours. Improving early detection is a major goal of current research: the two main approaches to achieving this goal are:

1. Development of a non-endoscopic screening tool.
2. Improvement of detection of dysplasia within current surveillance programmes to allow endotherapy and prevent progression.

This research project addresses both of these aims using emerging techniques from vibrational spectroscopy. The use of infrared spectroscopy to classify oesophageal pathology is investigated, with the goal of applying infrared spectroscopy to oesophageal cells collected using a non-endoscopic cell collection device. This would allow screening amongst the general population, and revolutionise models of surveillance for patients known to have Barrett's oesophagus.

Raman spectroscopy is investigated with the goal of developing a Raman probe designed for use within the instrument channel of a conventional endoscope. This technique could potentially improve endoscopic detection of dysplasia, and enable targeted endotherapy.

Both infrared and Raman spectroscopy are investigated as tools for automated histopathology: this could improve detection of dysplasia and reduce the workload of pathology departments in processing samples for Barrett's

surveillance. More specific objectives are discussed below in the relevant chapters.

SECTION A: LITERATURE REVIEW

CHAPTER 1: BARRETT'S OESOPHAGUS

1. Barrett's oesophagus

1.1. Definitions

Barrett's oesophagus owes its eponym to Norman Barrett, a British thoracic surgeon who first described the condition in 1950 (Barrett 1950). He proposed that the condition resulted from a congenitally short oesophagus, with the stomach extending into the mediastinum (Barrett 1950). It was subsequently recognised that the condition was in fact acquired, with an oesophageal origin for the columnar mucosa, and an association with gastro-oesophageal reflux (Allison and Johnstone 1953). Barrett himself accepted this view, and later referred to it as 'lower oesophagus lined by columnar epithelium' (Barrett 1957).

Barrett's is now well-recognised as representing a metaplastic change from the normal squamous epithelial lining of the oesophagus to a columnar epithelium, predisposing to an increased risk of oesophageal adenocarcinoma, though there is variation among expert opinion on the precise defining characteristics. The British Society of Gastroenterology (BSG) guidelines include both endoscopic and histological criteria for diagnosis:

'Barrett's oesophagus is defined as an oesophagus in which any portion of the normal distal squamous epithelial lining has been replaced by metaplastic columnar epithelium, which is clearly visible endoscopically (≥1cm) above the gastro-oesophageal junction and confirmed histopathologically from oesophageal biopsies' (Fitzgerald et al. 2014).

Meanwhile for the American Gastroenterological Association the only criterion is the presence of intestinal metaplasia (IM) on biopsy, whilst the American College of Gastroenterology also requires IM and an endoscopically visible segment of columnar mucosa (but do not have a minimum length requirement in their definition) (Spechler et al. 2011; Wang and Sampliner 2008).

The clinical significance of intestinal metaplasia is that it is the only histological subtype with strong evidence for risk of malignant progression – this is discussed in further detail, along with the other histological subtypes of columnar mucosa, in the section on histology (Fitzgerald et al. 2014; Spechler et al. 2011; Wang and Sampliner 2008).

1.2. Epidemiology of oesophageal adenocarcinoma

The incidence of oesophageal adenocarcinoma has shown a dramatic rise amongst Western populations during the latter portion of the 20th century (Edgren et al. 2013; Pera et al. 2005; Powell et al. 2002; Trivers, Sabatino, and Stewart 2008). Population data from the US and the UK showed a marked rise beginning in the 1960-70s (Powell and McConkey 1990; Yang and Davis 1988); this trend has continued and similar rises have been seen from Western Europe and Australia (Edgren et al. 2013). To date, the underlying cause of this rapid rise in incidence has not been fully explained (Edgren et al. 2013; Pera et al. 2005). Whilst strong associations with gastro-oesophageal reflux and obesity have been demonstrated, both of which have increased in Western populations in the late 20th century, these factors do not fully account for the changing epidemiology of this disease either alone or in combination, as population databases suggest an inflection point which precedes the increase in these risk

factors (Edgren et al. 2013; Pera et al. 2005). The rise in oesophageal adenocarcinoma also parallels the decreasing prevalence of Helicobacter Pylori infection in developed countries, suggesting the bacterium may have a protective effect (Labenz and Malfertheiner 1997). Data from the US suggest that the rate of increase has slowed, and incidence is reaching a plateau: national data show a yearly rate of increase of 8.2% from 1973 to 1996, followed by a 1.3% rise from 1996 to 2006 (Pohl, Sirovich, and Welch 2010).

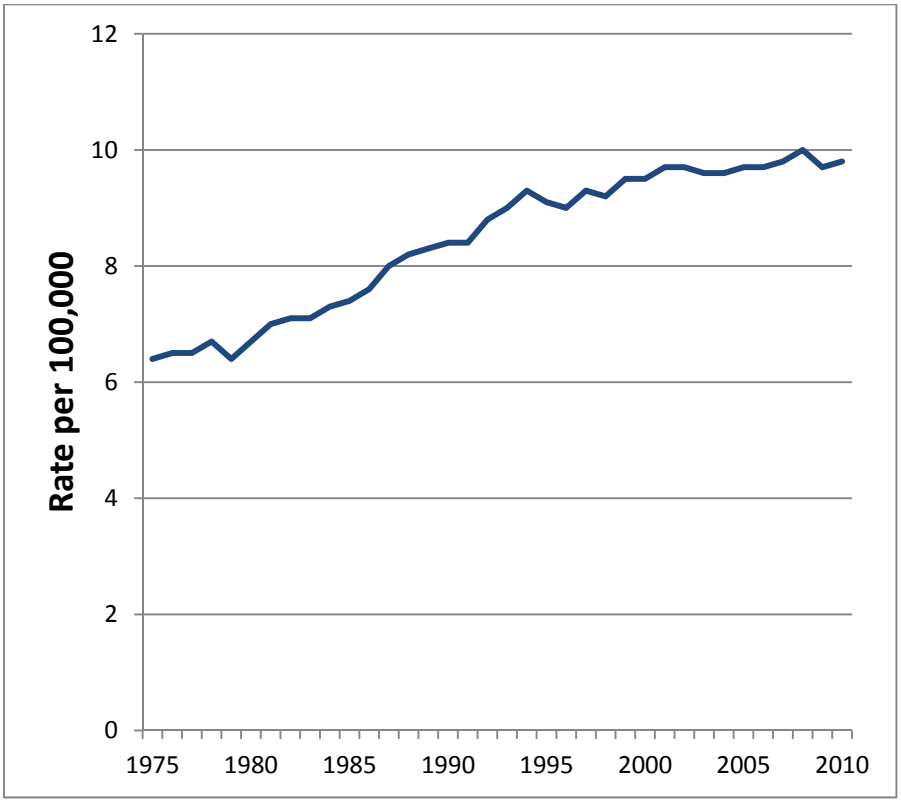


Figure 1-1 UK incidence of oesophageal cancer (all types), 1975-2010 Reproduced from Cancer Research: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/oesophagus/survival/#Trends (Accessed May 2014).

There are now over 8000 cases of oesophageal cancer in the UK each year, but survival rates remain poor: only 20-30% of oesophageal cancers are potentially curable at presentation, and overall 5-year survival rate in the UK 2005-9 was 13% (Cancer Research UK 2013).

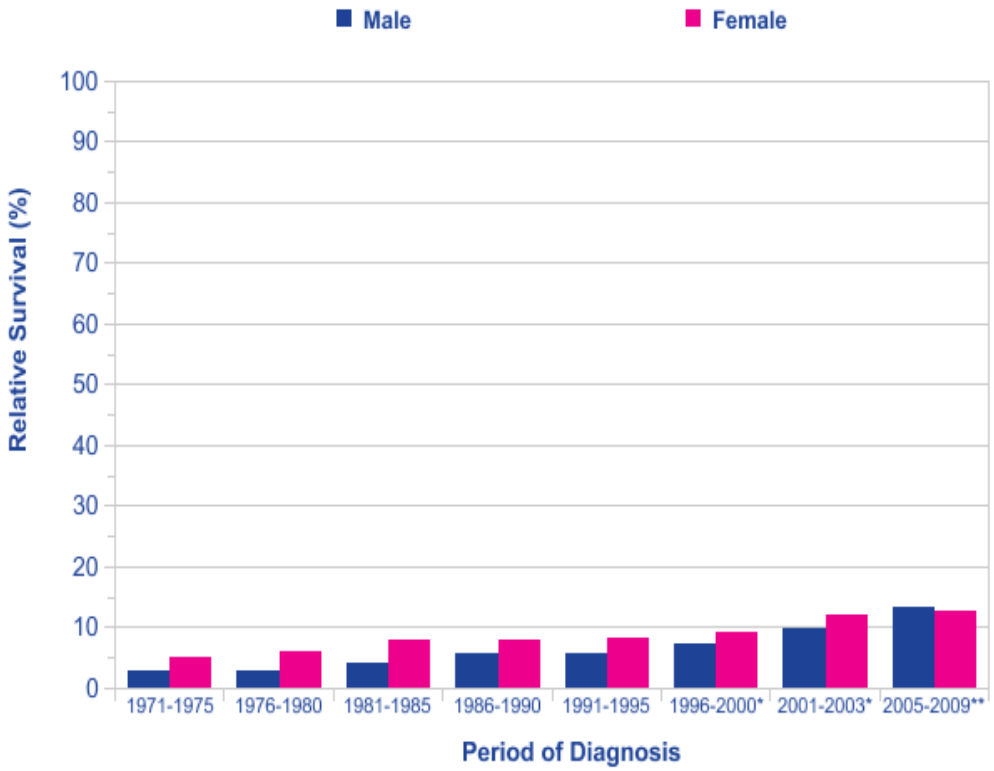


Figure 1-2 UK survival from oesophageal cancer (all types) over time, 1971-2009. Reproduced from Cancer Research: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/oesophagus/survival/#Trends (Accessed May 2014).

1.3. Epidemiology of Barrett's oesophagus

The majority of patients with Barrett's oesophagus (BO) are asymptomatic and remain undiagnosed, and thus the true population prevalence is uncertain (Cameron et al. 1990; Dulai et al. 2002; de Jonge et al. 2013). The population who become symptomatic and present may represent a subpopulation which could further confound epidemiological data (de Jonge et al. 2013).

Estimates have been derived from population studies, autopsy findings and studies undertaking endoscopy in volunteer populations (both symptomatic and asymptomatic groups).

Rex et al. (2003) offered upper gastrointestinal (GI) endoscopy to patients scheduled for colonoscopy with no prior history of oesophagogastroduodenoscopy (OGD) and found 6.8% overall had Barrett's oesophagus. Heartburn symptoms did not significantly increase the risk of Barrett's, but were associated with an increased risk of long-segment Barrett's (Rex et al. 2003). In a similar study in a group of patients all over the age of 65, an even higher prevalence of 16.7% was reported (Ward et al. 2006). However it is likely that the selection bias from including older patients with pre-existing indications for colonoscopy gives a significant over-estimate.

Population based studies in unselected patients have reported lower prevalences of BO, with a study in 1000 volunteers in Sweden finding a prevalence of 1.6%, and subsequent studies in Italy and China 1.3% and 1.9% respectively (Ronkainen et al. 2005; Zagari et al. 2008; Zou et al. 2011).

Extrapolating these estimates from unselected patients to the UK population would suggest there are in the region of 1 million people with Barrett's oesophagus in the UK.

1.4. Risk factors for Barrett's oesophagus

Gastro-oesophageal reflux disease (GORD) is the key risk factor for developing BO (Lagergren et al. 1999) and as such BO can be viewed as a response to chronic injury: a protective effect from frequent acid reflux. A meta-analysis of

26 studies found GORD to be associated with long segment BO, but the association for short segment BO (SSBO) was less clear (Taylor and Rubenstein 2010). In one cohort, 25% of patients undergoing endoscopy for symptoms of gastro-oesophageal reflux were found to have endoscopic columnar-lined oesophagus (Balasubramanian et al. 2012). This study also found heartburn duration >5 years, and hiatus hernia to be risk factors for BO (Balasubramanian et al. 2012). Hiatus hernia is often associated with symptoms of GORD, and has also been identified as a risk factor for BO in an earlier study (Avidan et al. 2002). A number of other conditions are less-commonly associated with a risk of BO and are shown in Table 1-1 below.

Table 1-1 Conditions associated with the development of Barrett's oesophagus

Chronic oesophageal reflux (>60% of cases)
Congenital retardation syndromes (1%)
Non-steroidal anti-inflammatory drugs (1%)
Chemotherapy (<1%)
Viral oesophagitis (<1%)

Reproduced from Jankowski et al. 2010.

Male gender and older age are the major demographic factors which predispose to BO, and Caucasian race has also been implicated (Balasubramanian et al. 2012; Cook, Wild, and Forman 2005; Edelstein et al. 2009; Eloubeidi and Provenzale 2001; Gerson et al. 2001). Central obesity appears to be a risk factor (both waist-to-hip ratio and abdominal

circumference) (Corley et al. 2007; Edelstein et al. 2007, 2009), but studies based on BMI have led to conflicting results (Anderson et al. 2007; Corley et al. 2007; Edelstein et al. 2007; El-Serag et al. 2005). A number of studies have implicated smoking as a risk factor for Barrett's but this has not been demonstrated in every study (Anderson et al. 2007; Edelstein et al. 2009; Kubo et al. 2009; Smith et al. 2005).

A recent study by Liu et al. attempted to risk stratify those referred for upper gastrointestinal endoscopy on demographic risk factors for BO (Liu et al. 2014). Their training model included 1,603 patients and determined the following factors predictive of BO (any length of columnar lined oesophagus): age, male sex, acid reflux (classified as 'never or sometimes' or 'often and daily'), and anti-reflux medication. Among this cohort of patients referred for endoscopy, abdominal pain and chest pain were negative predictors of BO. Smoking was not an independent risk factor once other risk factors were accounted.

As mentioned earlier, there is epidemiological evidence that Helicobacter pylori (H. pylori) infection may have a protective effect against oesophageal cancer (Labenz and Malfertheiner 1997). It has been posited that long term H. pylori infection leads to multifocal atrophic gastritis, and consequently reduced gastric acid secretion, as a possible mechanism of protecting against Barrett's oesophagus, and adenocarcinoma (Molloy and Sonnenberg 1997).

A positive family history for HGD or OAC increases the risk of developing BO, with one study calculating a 28% risk for first-degree relatives (Juhasz et al. 2011). Much recent work has focused on identifying genetic markers that are associated with an increased risk of BO and OAC. In a genome-wide association study comparing 1852 patients with Barrett's against 5172 controls

run by the Esophageal Adenocarcinoma Genetics Linkage consortium (EAGLE), 2 genetic loci were identified that were strongly associated with a small increased risk of Barrett's (6p21 and 16q24) (Su et al. 2012).

1.5. Natural history

Whilst the sequence of progression from metaplasia to low grade, then high grade dysplasia, followed by carcinoma, is well characterised (Jankowski et al. 1999), the exact risk of progression has been the subject of debate. The historically quoted cancer risk of 0.5% per patient per year was derived largely from small case series (Fitzgerald et al. 2014): more recent, population-based studies have revised this figure down. A study in Northern Ireland followed 8,522 patients with Barrett's oesophagus for a mean 7 years, and found risk of progression to high grade dysplasia (HGD) or adenocarcinoma was 0.22% per year, with a risk of 0.16% per year for adenocarcinoma alone (Bhat et al. 2011). A Danish study which identified patients with IM-type Barrett's oesophagus from the Danish Pathology Registry followed up 11,028 patients for median 5.2 years (Hvid-Jensen and Pedersen 2011). This group found progression to HGD or adenocarcinoma in 0.26% per year, and 0.12% for adenocarcinoma alone. Examining non-dysplastic Barrett's in isolation, the risk of adenocarcinoma was 0.10% per year.

A similar population-wide study carried out in the Netherlands followed up 42,207 patients with a histological diagnosis of Barrett's oesophagus, of whom 16,333 underwent surveillance endoscopy (though the clinical reasons for choosing surveillance were not recorded in this retrospective study) (de Jonge et al. 2010). In those who entered surveillance, after excluding patients with a

diagnosis of HGD or adenocarcinoma within 12 months of the first endoscopy, the risk of progression to HGD or adenocarcinoma was 0.58%, and 0.43% for adenocarcinoma alone. Examining the group as a whole and including those who did not receive surveillance endoscopy, the risk of progression to adenocarcinoma was lower at 0.14%.

A large meta-analysis totalling 58,547 patient-years of follow up reported an incidence of adenocarcinoma of 0.33% per patient per year (Desai et al. 2012). However there is significant geographical variation in incidence of oesophageal adenocarcinoma, and it is possible that the risk for patients with Barrett's varies between regions. One meta-analysis of data exclusively from the UK (prior to the study by Bhat et al. in 2011) suggested the risk for UK patients is around 1.0% (Jankowski, Provenzale, and Moayyedi 2002).

One study suggested that the risk of progression to adenocarcinoma reduces with length of time in surveillance (Gaddam et al. 2013). Stratifying patients by number of endoscopies under surveillance (and hence time), those patients with persistent Barrett's but no progression to dysplasia had a reducing risk of progression with time under surveillance. This suggests that those who are going to progress are more likely to do so early in the course of disease, and the authors conclude that endoscopy intervals could be lengthened for patients under surveillance for several years with stable pathology.

There is further variation in the reported incidence of progression of low grade dysplasia (LGD). The Danish group found the risk of progression to adenocarcinoma if LGD was detected at index endoscopy was 1.27% per year (Hvid-Jensen and Pedersen 2011). Similarly, a study of 210 patients with LGD followed up for mean 6.2 years found a combined risk of HGD/oesophageal

adenocarcinoma (HGD/OAC) of 1.86% per year (Wani et al. 2011). However this study also noted significant inter-observer disagreement on diagnosis of LGD even by experienced pathologists. Other studies have suggested a higher rate of progression. A large multi-centre cohort study followed 156 patients with LGD for a mean 4.1 years and found progression to HGD/OAC in 13% (Sharma, Falk, et al. 2006). A recent RCT of radiofrequency ablation (RFA) versus surveillance for LGD reported a rate of progression to HGD/OAC of 26.5% over 3 years in the control arm, with progression to OAC in 8.8% (Phoa et al. 2014).

The variation in reported risk may reflect the difficulties in achieving consensus pathology with LGD: over diagnosis may result in underestimates for the true risk of progression.

For patients diagnosed with HGD on biopsies there is a significant risk that intra-mucosal adenocarcinoma may be present but has not been identified on the initial biopsies. A number of studies have examined oesophagectomy specimens performed for HGD. A systematic review of these studies calculated a pooled average of 39.9% for the occurrence of either intra-mucosal or invasive OAC (Konda et al. 2008). The risk of invasive OAC was 11% in those with visible lesions, and 3% for those with no visible lesion.

1.6. Diagnosis

1.6.1. Histopathology

1.6.1.1. The normal oesophagus: squamous epithelium

In cross section the oesophageal wall is composed of mucosa, which is separated from submucosa by the muscularis mucosae, with circular and longitudinal muscle layers outside the submucosa, all surrounded by a thick covering of connective tissue, the adventitia (Shepherd 2013).

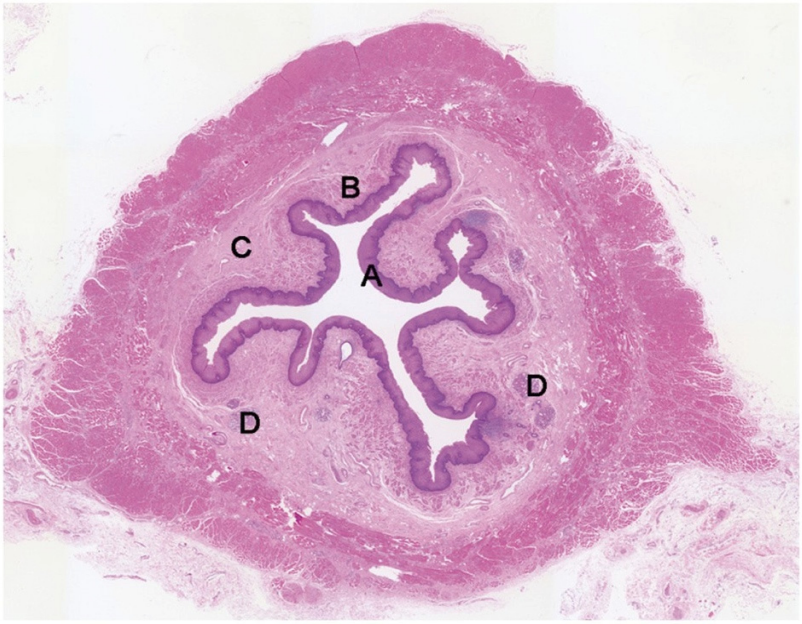


Figure 1-3 Microanatomy of the wall of the oesophagus. A Normal squamous epithelium-lined mucosa; B Muscularis mucosae; C Submucosa; D Submucosal glands. Figure reproduced from Shepherd 2013.

The mucosa consists of a non-keratinising stratified squamous epithelial lining, approximately 500-800μm thick (Shepherd 2013). This lies on a lamina propria, a loose connective tissue layer which separates the mucosa from a layer of muscle fibres, the muscularis mucosae. The submucosa contains gland structures, nerves and lymphatic vessels within a loose connective tissue

network. This rich lymphatic plexus contributes to the high rates of lymph node metastasis with oesophageal cancers invading the submucosa.

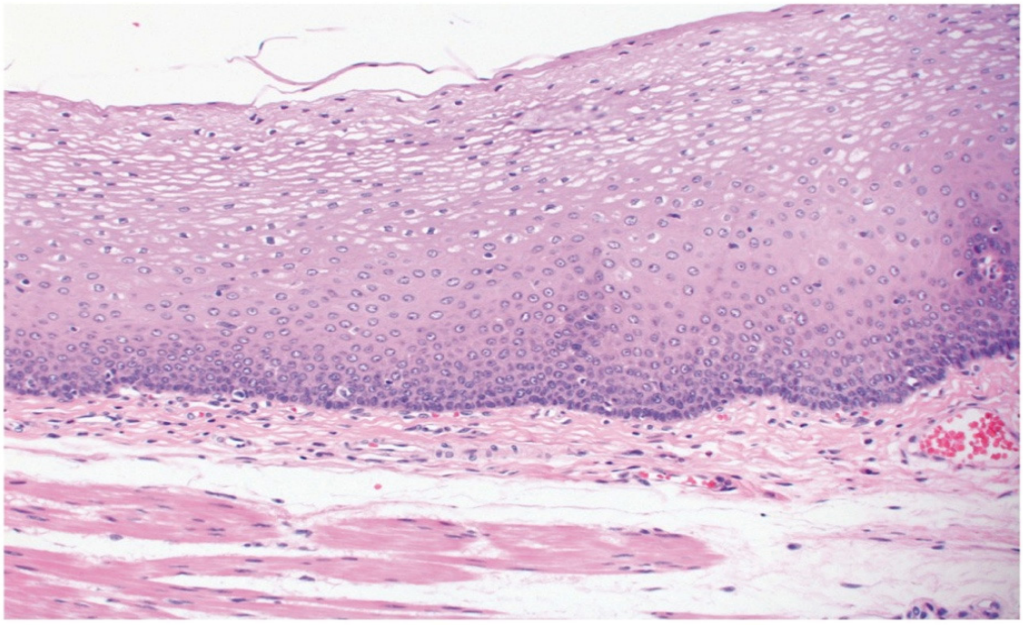


Figure 1-4 Normal squamous epithelium of the oesophagus. Figure reproduced from Shepherd 2013.

The extracellular matrix is composed of structural proteins such as collagen and fibronectin, and enzymes including the matrix metalloproteinases. This microenvironment can play a role in the dysplastic progression of Barrett's oesophagus, provide support for developing tumours and promote tumour invasion (Picardo et al. 2012). Changes to its composition may occur with neoplasia and a number of matricellular proteins have been identified with altered expression in Barrett's oesophagus or adenocarcinoma. Secreted protein acidic and rich in cysteine (SPARC) has multiple effects such as cell cycle modulation and antiproliferative functions, and is increased in BO and OAC (Botelho et al. 2010). Thrombospondin-1 (TSP-1) has an anti-angiogenic action and regulates growth factors, and is differentially expressed in OAC (Saadi et al. 2010). Osteopontin is another matricellular protein known to

promote metastasis that has been found to upregulated in OAC (Miller et al. 2006). The matrix metalloproteinase (MMP) enzymes have a major role in turnover and remodelling of the extracellular matrix, along with tissue growth and angiogenesis. A number of MMPs (particularly MMP-1, -2, -3, -7 and -9) have been found to be increased along the progression from BO to OAC (Grimm et al. 2010; Groblewska et al. 2012; Herszenyi et al. 2007; Keld et al. 2010; Lagarde et al. 2007; Miller et al. 2006; Salmela et al. 2001), suggesting that these changes occur early in the metaplasia-dysplasia-adenocarcinoma sequence. A study by Lagarde et al. (2007) identified MMP-1, -3, -7 and -9 as prognostic biomarkers for OAC.

1.6.1.2. Barrett's oesophagus

As discussed earlier, BO is the replacement of the normal squamous epithelium with columnar epithelium through metaplastic change. This is thought to be chiefly a response to gastro-oesophageal reflux, with both the high acid content, and the reflux of bile and alkali from the duodenum thought to play a part in the development of neoplasia (Dixon et al. 2001; Lagergren et al. 1999; Vaezi and Richter 1996). Using a canine model, replacement of squamous epithelium with columnar epithelium was seen following high acid exposure (Gillen et al. 1988). This was demonstrated in the presence of squamous barriers to proximal migration of gastric columnar epithelium, thus highlighting that alternative mechanisms are involved other than simply proximal extension of gastric epithelium. A further argument against gastric migration has been the absence of goblet cells in non-metaplastic gastric epithelium. Other proposed candidates include pluripotent basal stem cells in squamous epithelium, or columnar cells

lining oesophageal gland ducts (Nicholson et al. 2012). However it has been recently argued that the evidence from non-animal models may be misleading, and the evidence from human studies points most strongly towards a cellular origin in the proximal columnar gastric or cardiac epithelium (McDonald et al. 2015a). It thus seems most likely that development of Barrett's oesophagus is a two-stage process, with migration of gastric epithelium followed by metaplastic change to intestinalisation with goblet cells (McDonald et al. 2015a;McDonald et al. 2015b).

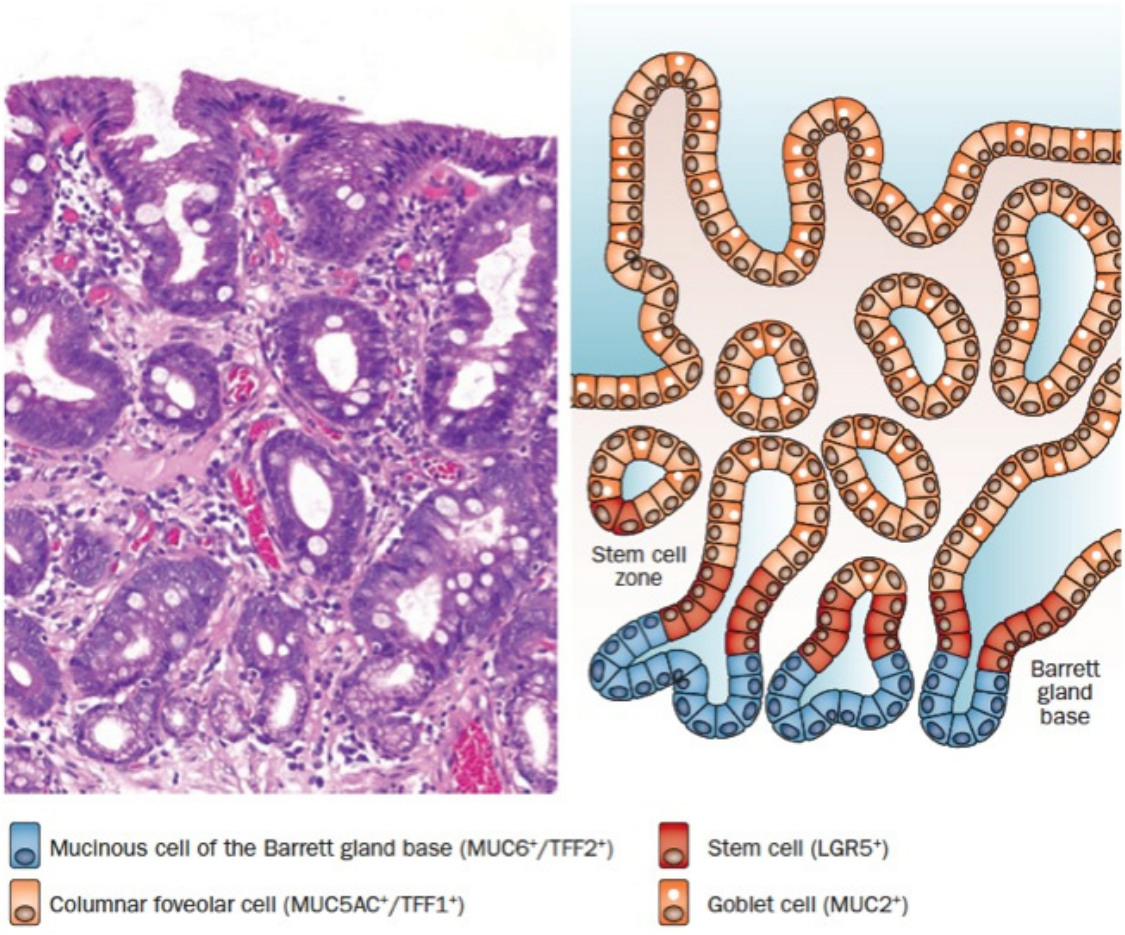


Figure 1-5 The structure of a typical Barrett's gland. Haematoxylin and eosin shown in A, with corresponding protein expression highlighted in the schematic diagram in B. Reproduced from McDonald et al. 2015b.

The stem cells which form the metaplastic glandular epithelium in Barrett's arise in the neck of the gland and exhibit bidirectional migration (Lavery et al. 2014). Recent research has shown clonality of metaplastic glands in Barrett's, indicating a common derivation from a single progenitor cell (Nicholson et al. 2012). Multiple differentiated cell lineages have demonstrated clonality, indicating that pluripotent stem cells can populate large areas of epithelium (Nicholson et al. 2012). In the work by Nicholson et al. (2012), one patient exhibited a common clonal mutation in both glandular metaplastic tissue and regenerating squamous epithelium, suggesting a common cellular origin and therefore indicating plasticity in the undifferentiated precursor cell.

Thus the current theory is that Barrett's glands are clonal units sustained by multiple stem cells derived from a single (multipotent) common progenitor (McDonald et al. 2015b).

There are three main types of epithelium that form columnar-lined oesophagus (CLO): cardiac, fundic and intestinal metaplasia (IM). A definitive histological diagnosis of CLO can only be made in the presence of native oesophageal structures such as submucosal glands or their associated ducts, or islands of squamous epithelium amongst the metaplastic glandular mucosa, which are invariably associated with submucosal glands (Coad et al. 2005). These features confirm that the columnar epithelium is unequivocally in the oesophagus, but are only present in around 10-15% of biopsy sets (Srivastava et al. 2007; Takubo, Nixon, and Jass 1995). Typically, a 'patchwork' of the three different subtypes of CLO is seen on biopsy samples (Shepherd 2013).

Cardiac and fundic-type metaplasias are so named because of their resemblance to the columnar epithelium that normally lines these two areas of

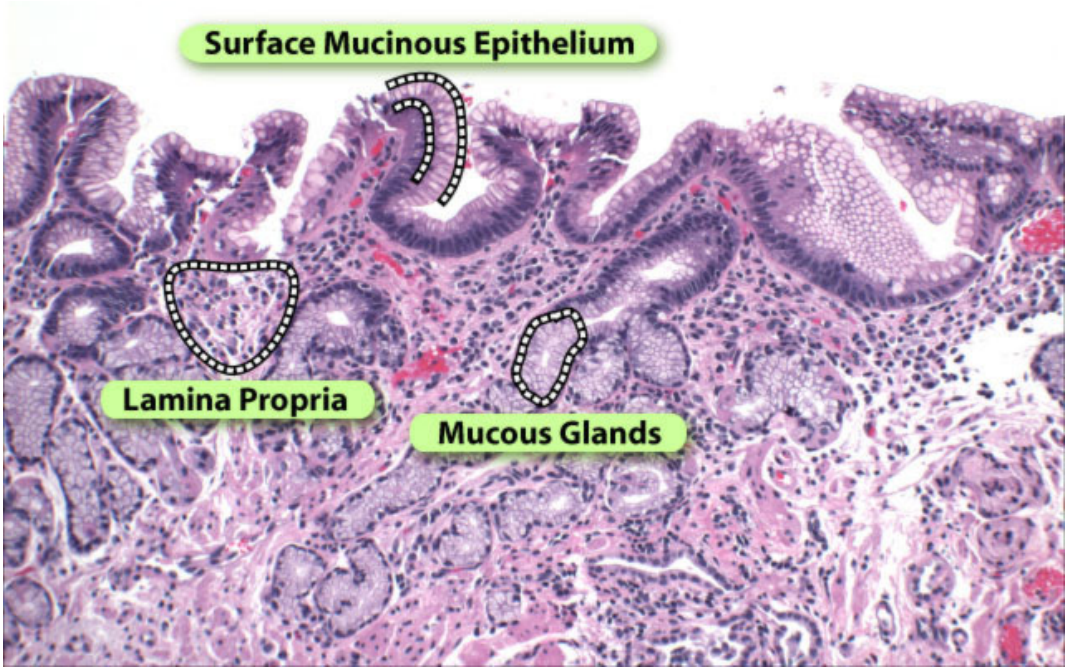


Figure 1-6 Cardiac-type mucosa. Figure reproduced from http://www.barrettsinfo.com/content/3b_what _is_histology.cfm, (accessed 20/01/15).

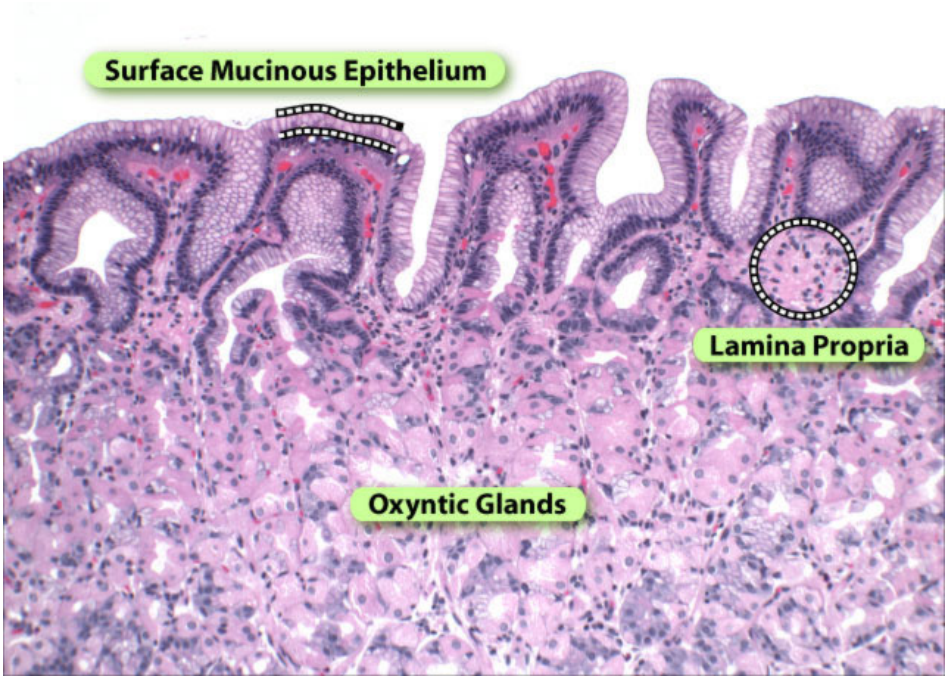


Figure 1-7 Fundic-type mucosa. Figure reproduced from http://www.barrettsinfo.com/content/3b_what _is_histology.cfm, (accessed 20/01/15).

the stomach (cardia and fundus). Cardiac-type metaplasia contains glands but no oxyntic cells or goblet cells. Fundic-type metaplasia contains oxyntic cells but no goblet cells. Correlation with the site of biopsy at endoscopy is essential for correct classification here.

The characteristic feature of intestinal metaplasia is the presence of goblet cells. Another feature which may be seen is non-goblet cells which stain positively with Alcian blue and indeed these 'blue cells' may be more prevalent than goblet cells (Shepherd 2013).

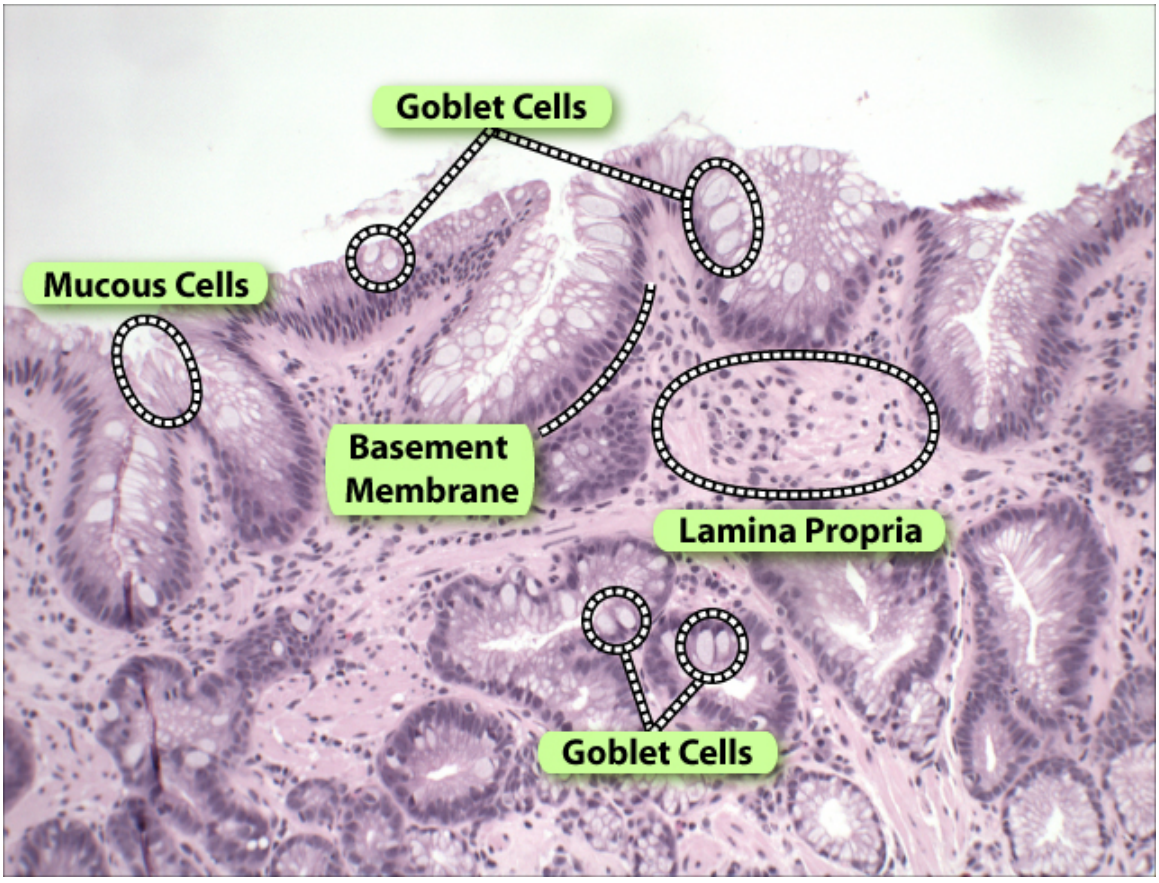


Figure 1-8 Intestinal metaplasia.

Figure reproduced from http://www.barrettsinfo.com/content/3b_what_is_histology.cfm, (accessed 20/01/15).

The role of IM in the diagnosis of BO is controversial. There is evidence to suggest that the risk of malignant progression associated with BO is much higher in those with identified IM than those with CLO without IM (Bhat et al. 2011; Skinner et al. 1983). Difficulty arises, firstly, in identifying the presence of IM due to sampling error from mucosal biopsies. One study looked at long-segment BO in 125 patients with a total of 1646 biopsies and concluded that, to demonstrate goblet cells in 67.9% of endoscopies required 8 biopsies, but if only 4 were taken goblet cells were identified in only 34.7% (Harrison et al. 2007). Thus the absence of IM on biopsy does not reliably preclude the presence of IM. A further study followed patients with (612 patients) and without IM (322 patients) at index biopsy (Gatenby et al. 2008). They found that over 50% of patients without IM were found to have IM at 5-year follow up, and over 90% had IM at 10 years, though the extent to which this represents sampling error or new development of IM is not clear. The rates of progression to dysplasia and cancer were similar between groups.

There is also emerging evidence that CLO that does not contain goblet cells may still show ‘intestinalisation’, and harbour molecular changes that suggest potential for malignant progression (Chaves et al. 2007; Hahn et al. 2009; Liu et al. 2009; Romagnoli et al. 2001).

Despite these difficulties, the current US management guidelines all require IM as a necessary component of BO diagnosis (Spechler et al. 2011; Wang and Sampliner 2008). Historically, pathological opinion in the UK has successfully argued that IM should not be a defining feature of BO (Shepherd 2013; Watson, Heading, and Shepherd 2005), but this argument continues (Riddell and Odze 2009). The most recent BSG guidelines have taken an intermediate position in

which IM is not required for the diagnosis of BO, but patients with biopsy-proven IM are managed differently from those without (discussed in section 1.9 below) (Fitzgerald et al. 2014). Commenting on the presence of IM is therefore a key component of the pathological reporting terms advocated in the updated BSG guidelines (see Table 1-2).

Table 1-2 Pathological reporting of specimens

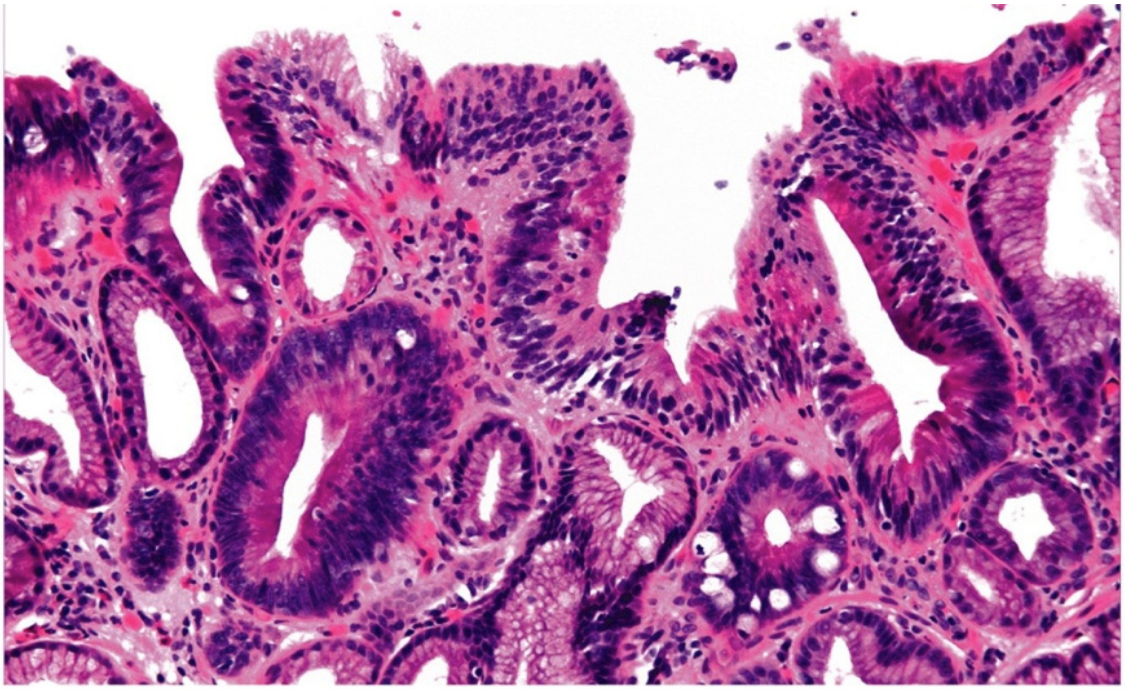
'Barrett's oesophagus with gastric metaplasia only' (glandular epithelium with cardiac/fundic metaplasia)
'Barrett's oesophagus with IM'
'No evidence of Barrett's oesophagus' (squamous epithelium without glandular tissue)

Adapted from BSG guidelines (Fitzgerald et al. 2014).

1.6.1.3. Dysplasia

Dysplasia in BO is classified using the Riddell-type classification, as 'indefinite for dysplasia', low grade dysplasia (LGD), and high-grade dysplasia (HGD) (Shepherd 2013). The Vienna classification attempted to standardise diagnostic terminology worldwide, and resolve some of the differences between Western and Japanese classifications (Schlemper et al. 2000). Although both cytological and architectural features of dysplasia are well-described, there are no definitive criteria for dysplasia, and the distinction between grades of dysplasia is even more blurred. Some examples of LGD and HGD are shown in Figure 1-9.

A



B

Figure 1-9 Examples of A) low grade dysplasia and B) high grade dysplasia. Figures reproduced from Shepherd 2013.

The diagnosis of LGD is made largely on the basis of cytological atypia, with relative preservation of tissue architecture. HGD generally has abnormal glandular architecture and more marked cytological atypia (Shepherd 2013). A number of features are described that can help differentiate the two, and these are presented in Tables 1-3 and 1-4 below.

Table 1-3 Pathological characteristics of low and high grade dysplasia

Low Grade Dysplasia	**High Grade Dysplasia**
Cytology approximates to that of mild and moderate adenomatous dysplasia	Cytology approximates to that of severe adenomatous dysplasia
Nuclei enlarged, crowded, hyperchromatic and ovoid	Nuclei enlarged, usually spheroidal, may have open chromatin pattern with nucleoli
Mitotic activity may be substantial and atypical mitoses may be present	Mitotic activity may be substantial and atypical mitoses usually present
Stratification is often present	Stratification may be present but there is usually pronounced cellular disorganization
Architectural change, including villosity, may be present but in the appropriate cytological setting there is loss of the basal-luminal maturation/differentiation axis	Architectural change, including villosity, glandular budding and complex glandular structures, is often present. There is loss of the basal-luminal maturation/differentiation axis

Table adapted from Fléjou and Svrcek 2007 in Shepherd 2013.

It is important to appreciate that the distinctions between LGD and HGD are arbitrary cut-offs along a biological continuum, and as such strict definitions are not possible. Difficulties with poor inter-observer reliability in classifying dysplasia are well recognised (Kaye et al. 2009; Kerkhof et al. 2007; Montgomery et al. 2001; Reid et al. 1988; Shepherd 2013). This is particularly true for indefinite for dysplasia, and LGD (Shepherd 2013). A number of studies have shown that the prediction of progression is more accurate with consensus pathology (Curvers, ten Kate, et al. 2010; Skacel et al. 2002), and this has led to the current guidance for all dysplastic specimens in BO to be reported by a second gastrointestinal pathologist (Fitzgerald et al. 2014).

Immunohistochemical biomarkers can be used by pathologists to provide further information for diagnosis. The most widely studied in BO is p53. The ability of p53 to stain positively for dysplasia has been variably reported (Kaye et al. 2009, 2010; Khan et al. 1998; Murray et al. 2006) but there is evidence that it may improve interobserver agreement between pathologists (Kaye et al. 2009; Skacel et al. 2002) and serve as a marker of risk of progression (Kastelein et al. 2013; Murray et al. 2006). The current BSG guidelines advise that it should be considered as an adjunct to improve diagnostic reliability, but some authors still consider its use to be experimental (Fitzgerald et al. 2014; Shepherd 2013).

Table 1-4 Comparison of cytological and architectural features of low and high grade dysplasia

Feature	LGD	HGD
Cytology		
Increased nuclear:cytoplasm ratio	+	++
Loss of cell polarity	-	+
Mitosis	+	++
Atypical mitosis	+/-	+
Full-thickness nuclear stratification	-	+
Decreased goblet cells (+/- dystrophic)	+	++
Hyperchromasia	+	++
Multiple nucleoli	+/-	+/-
Large irregular nucleoli	-	+/-
Irregular nuclear contour	+	++
Nuclear pleomorphism	-	+
Architecture		
Villiform change	-	+/-
Crypt budding/branching	+/-	++
Crowded (back-to-back) crypts	+/-	++
Irregular crypt shapes	+/-	+
Intraluminal papilla/ridges	-	+/-
Lamina propria between glands	+	+/-

Adapted from Odze 2006 in Shepherd 2013.

1.6.1.4. Indefinite for dysplasia

There are several reasons why it may not be possible for the pathologist to give a certain diagnosis of dysplasia based on the biopsy samples obtained from a patient. Significantly, this term may be used both for high quality samples where there is a low degree of suspicion of LGD, but may also be used for samples of low quality that contain possible features of dysplasia, which could include HGD. Clarification by the pathologist when using this term is helpful to guide further patient management.

Difficulties in interpretation may be caused by technical factors such as crush artefact, poor staining or tangential sectioning (Shepherd 2013). Alternatively, other pathology or treatment artefacts can alter appearances, e.g. inflammatory and regenerative change can be hard to differentiate from dysplasia, so areas of oesophagitis or ulceration may preclude a confident diagnosis of dysplasia.

'Surface maturation' describes the loss of atypia seen as cells migrate from gland bases to the surface, and is recognised as a feature differentiating normal, regenerating mucosa from true dysplasia (Shepherd 2013). Cytological atypia in basal crypts is not an uncommon finding, having been reported in up to 7.3% of cases of BO (Lomo et al. 2006). However, some argue for a distinct case in which a focus of low/high-grade cytological atypia in a crypt stands out as markedly different from surrounding crypts, and that this 'crypt dysplasia' (or 'basal crypt dysplasia-like atypia') is a subtype of true dysplasia (Coco et al. 2011).The argument is that this represents an early stage in the development of dysplasia. There is controversy over the existence/significance of 'crypt dysplasia', with some authors remaining to be convinced that this is a form of

dysplasia (Shepherd 2013). Consequently this may be reported as LGD/HGD depending on the degree of dysplasia present, or 'indefinite for dysplasia' by those who do not recognise it as a form of dysplasia, and who argue that only this labelling will lead to the correct management for patients (Fitzgerald et al. 2014; Shepherd 2013).

1.6.1.5. Oesophageal adenocarcinoma

Once there is evidence of invasion beyond the basement membrane this signifies a carcinoma.

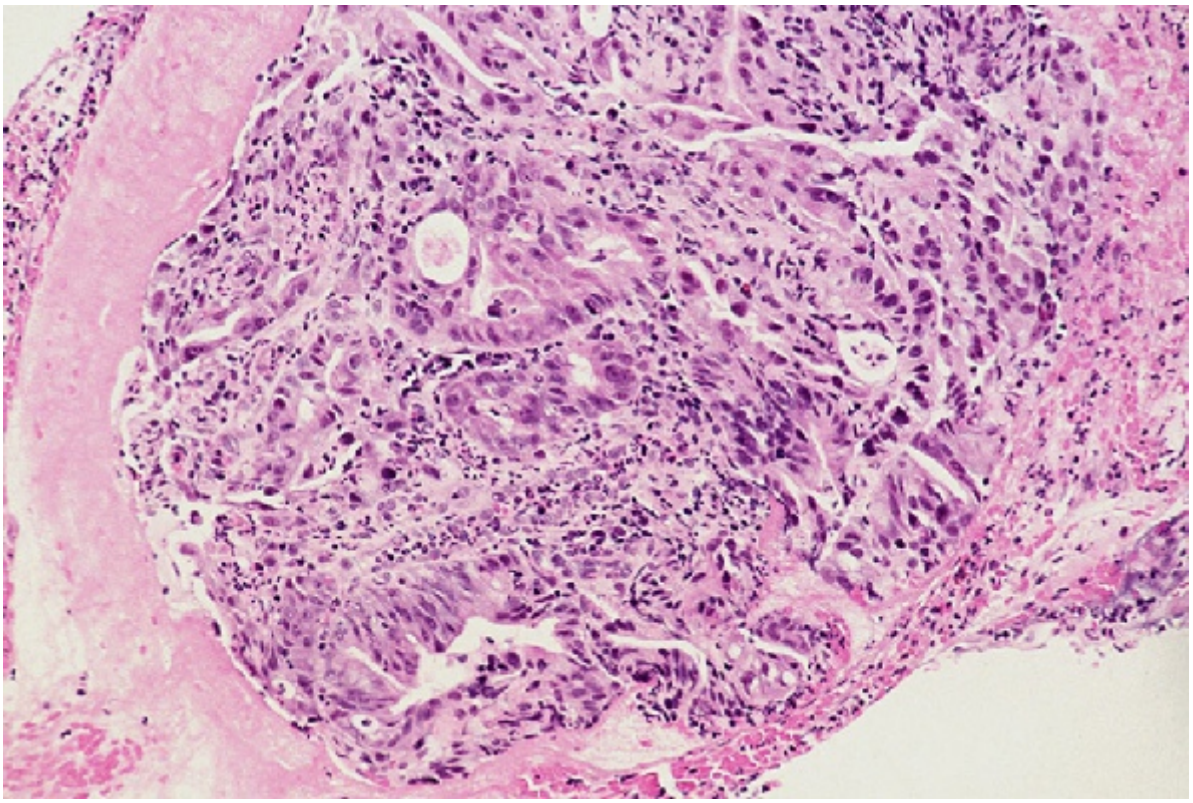


Figure 1-10 Oesophageal adenocarcinoma: intramucosal carcinoma. Figure reproduced from http://pathology2.jhu.edu/beweb/fig7.htm, accessed 20/01/15).

The tumour-node-metastasis (TNM) staging system is the most widely used classification system for oesophageal cancers, which reports the stage in terms of primary tumour (T), nodal spread (N) and distant metastases (M) (Raimes 2009). A table outlining current TNM staging for oesophageal cancer is presented in Appendix I.

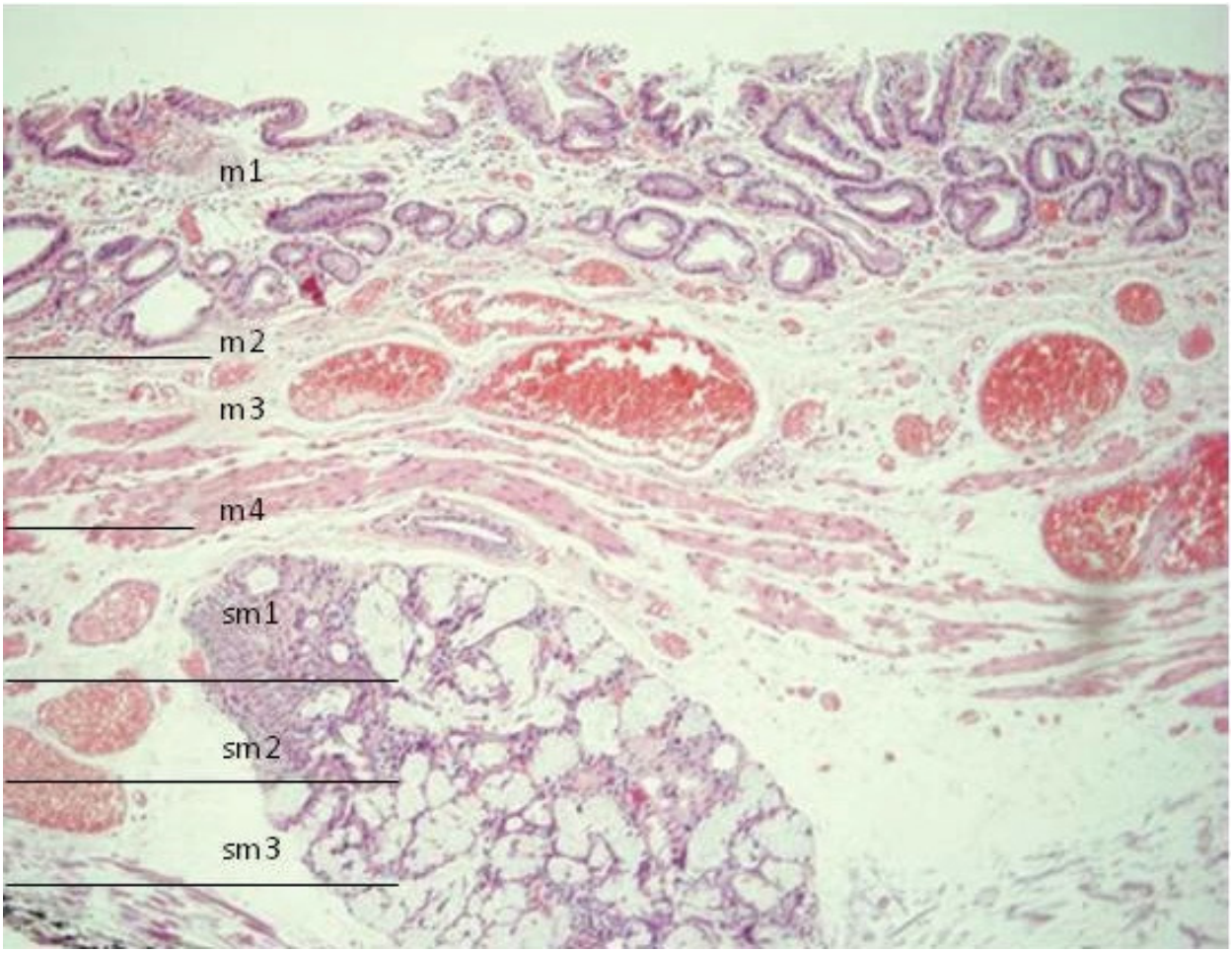


Figure 1-11 Classification of early oesophageal cancer in columnar lined oesophagus. Figure reproduced from Pech et al. 2008. Mucosal depth is shown as m1-m4, submucosal depth is shown as sm1-sm3.

In recognition of the clinical significance of invasion of certain layers of the submucosa, (discussed further in section 1.10 below), there is further subdivision for early cancers, and the following terms may be used in the context of early oesophageal cancer:

Intramucosal carcinoma Intramucosal carcinoma (IMC) is defined as penetration of the epithelial basement membrane into the lamina propria or muscularis mucosae, but without invasion of the submucosa.

Submucosal invasion by adenocarcinoma Any oesophageal adenocarcinoma that has invaded into or beyond the submucosa.

Classification is based on the depth of invasion of either the mucosa (m1-4) or submucosa (sm1-3) as shown in Figure 1-11.

1.6.2. Oesophageal cytology

Oesophageal cytology was performed relatively frequently in the era before flexible endoscopy as a means of diagnosing oesophageal cancer (Rubin 1960). Following the introduction of the flexible endoscope, some continued to advocate brush cytology as a complementary part of routine assessment for Barrett's surveillance, in addition to biopsy (Falk 2003). However, cytological assessment does not form part of the current Barrett's guidelines from the BSG (Fitzgerald et al. 2014), the AGA (Spechler et al. 2011) or the ACG (Wang and Sampliner 2008).

Diagnosis of Barrett's with intestinal metaplasia can be made on brush cytology by the recognition of goblet cells (Conrad, Cobb, and Raza 2012). Key features of adenocarcinoma cells are enlarged, pleomorphic nuclei, irregular nuclear

membranes and prominent nucleoli. The cytoplasm is variable in amount and may contain vacuoles. Cells appear in small overlapping clusters with loss of polarity (Conrad et al. 2012). High grade dysplasia appears similar to adenocarcinoma but without the tumour diathesis (necrotic debris, fibrin and inflammatory cells) and lacks the cellular dispersion and discohesion of adenocarcinoma (Conrad et al. 2012). Low grade dysplasia can be difficult to differentiate from reactive changes, which show nuclear enlargement and an inflammatory background (Conrad et al. 2012).

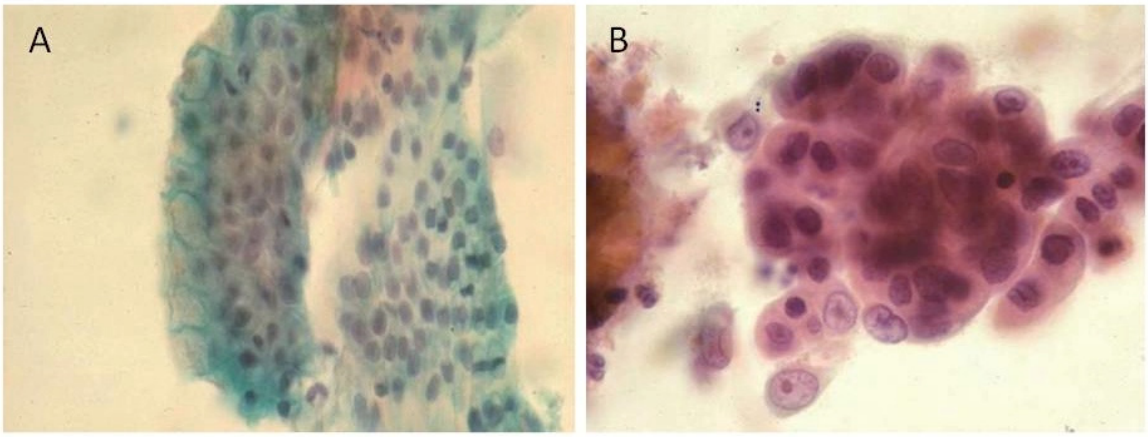


Figure 1-12 Examples of oesophageal cytology A) Barrett's oesophagus, glandular epithelium with goblet cells, B) adenocarcinoma with enlarged nuclei, prominent nucleoli and necrotic background. Figure adapted from Conrad et al. 2012.

Hughes and Cohen (1998) reviewed a number of studies comparing cytology and histology, and found the range of reported correlation between the two techniques in patients with dysplasia was 50-100% (Hughes and Cohen 1998). Of the larger studies which have compared brush cytology and histology among patients with known or suspected Barrett's neoplasia, Geisinger et al. found

agreement of 72% between the techniques (Geisinger, Teot, and Richter 1992), and Hardwick et al. reported sensitivity and specificity of 89% for brush cytology versus endoscopy and biopsy (Hardwick et al. 1997).

Oesophageal cytology can be difficult to interpret, particularly in the presence of inflammation or ulceration, and some authors advise that management decisions should not be made on the evidence of cytology alone as benign Barrett's in the presence of ulceration may appear malignant (Hardwick et al. 1997).

Non-endoscopic balloon cytology technique has been used to collect cells. This technique has been used as a screening tool in populations with a high prevalence of squamous cell carcinoma. This has been implemented in clinical practice in China for several decades (Dawsey et al. 1997; Shu 1983; Spechler 1997).

A study examining balloon cytology in Barrett's-associated neoplasia compared it against brush cytology, using endoscopy and biopsy as the gold standard, among a population of patients with suspected Barrett's oesophagus (Falk et al. 1997). Adequate collection of columnar epithelium was achieved in 52/63 cases (83%) for balloon cytology, versus 59 of 61 (97%) cases with brush cytology. Balloon cytology had sensitivity 80% for HGD and adenocarcinoma, but sensitivity of LGD was 25% (Falk et al. 1997). Brush cytology had sensitivity of 100% in HGD and adenocarcinoma but only 22% in LGD (Falk et al. 1997). Whilst these results imply that balloon cytology is not a suitable screening test for Barrett's neoplasia with this device, it suggests that brush cytology has the potential to collect a representative sample, but the difficulty of interpretation of

low grade dysplasia may limit the sensitivity using conventional cytological assessment.

1.6.3. Endoscopic diagnosis

White-light high-resolution endoscopy remains the gold standard for diagnosing Barrett's oesophagus. Accurate identification of the gastro-oesophageal junction (GOJ) is essential for diagnosis and measurement of BO. It is essential to differentiate an irregular squamocolumnar junction ('Z-line') from tongues of columnar-lined oesophagus, and recognise a hiatus hernia rather than mislabel this as BO. The squamocolumnar junction normally delineates the GOJ, but in the presence of Barrett's columnar mucosa will line the lower oesophagus. The 'diaphragmatic pinch' can also be helpful, but in the presence of a hiatus hernia the stomach may be in the thorax. Therefore the most reliable endoscopic indicator of the GOJ is the upper limit of the longitudinal gastric folds (Amano et al. 2006), but this can only be reliably assessed with sufficient air insufflation. Palisade vessels are present in the oesophagus and continue to the lower oesophageal sphincter (Shepherd 2013) and may be identifiable endoscopically. Some units, particularly in Japan, use the lower limit of palisade vessels to define the GOJ, but this has not shown good inter-observer agreement (Amano et al. 2006; Shepherd 2013).

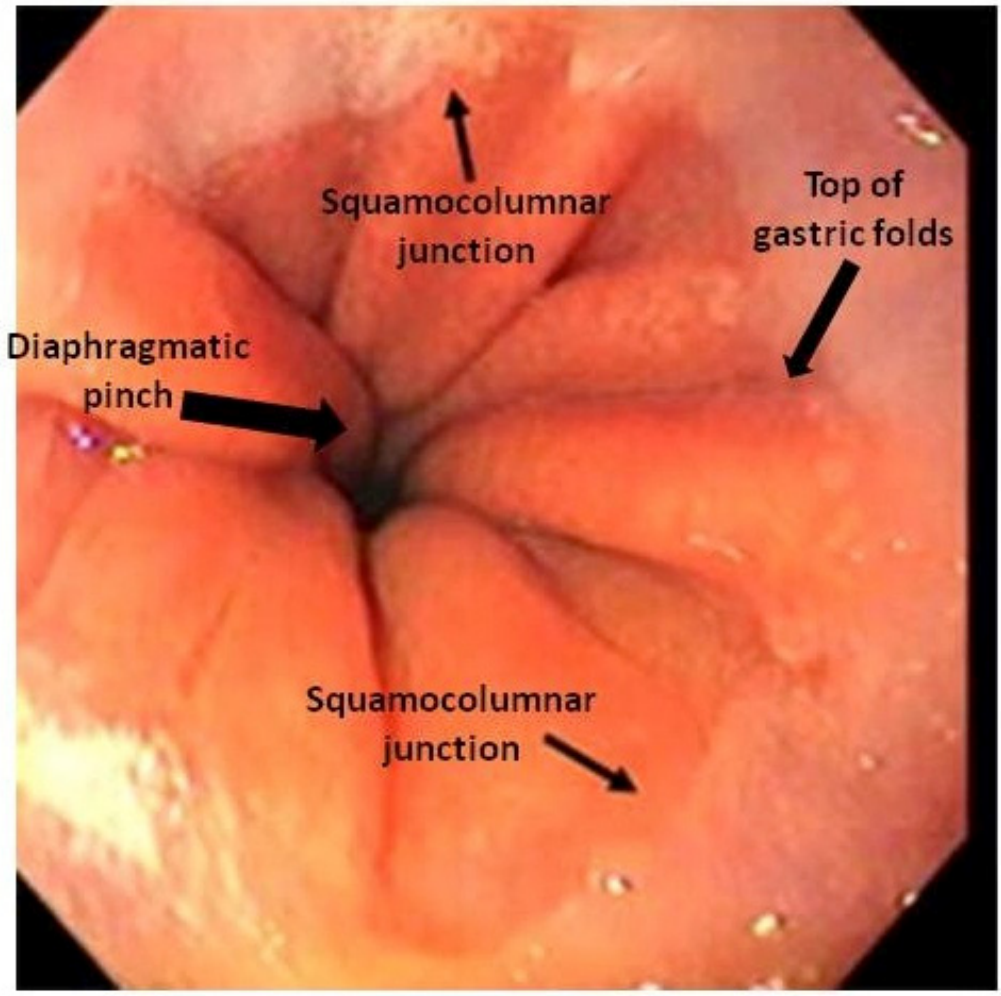


Figure 1-13 Identification of the gastro-oesophageal junction. File adapted from http://www.drleith.co.za/anti-reflux-surgery.php, accessed 20/01/15.

Endoscopic findings of BO should be reported in accordance with the Prague C&M classification, as agreed by the International Working Group for Classification of Oesophagitis (Sharma, Dent, et al. 2006). This advises that the length of the Barrett's segment is recorded as the circumferential (C) and maximal (M) lengths in centimetres. This classification system has been shown to have good inter-observer reliability for Barrett's lengths to within 2cm (rates of agreement for C and M 97% and 95% respectively), though exact agreement was less good (53% and 38% respectively) (Sharma, Dent, et al. 2006).

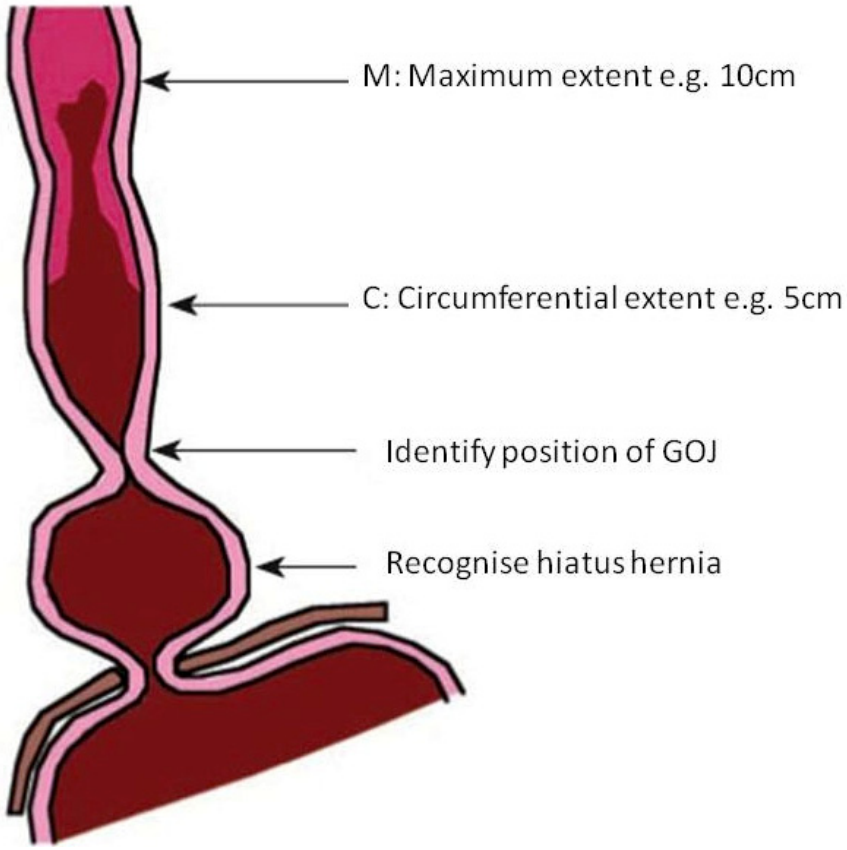


Figure 1-14 Reporting Barrett's length using Prague C&M criteria. Figure adapted from Nguyen-Tang, Frossard, and Dumonceau 2011.

Islands of BO should be described in terms of distance from the incisors and the length of the segment (Fitzgerald et al. 2014). The position of visible lesions should also be described in terms of their distance from the incisors, and their appearance using the Paris classification (see Table 1-5).

As discussed earlier, the diagnosis depends on both endoscopic and histological findings. In recognition of the need for confirmatory tissue diagnosis, along with the difficulties in accurately identifying dysplastic change at endoscopy, the Seattle biopsy protocol has been developed (Levine et al. 2000). This involves taking four-quadrant biopsies at 2cm intervals throughout the entire Barrett's segment, preferably with 'jumbo' biopsy forceps, in addition

to sampling any visible lesions. For long segments this can be a lengthy procedure: a 10cm circumferential segment of Barrett's requires 24 biopsies (assuming there are no visible lesions), and this number of specimens generates a significant workload for pathology services.

Table 1-5 Paris classification of superficial lesions (type '0' lesions) in the upper GI tract

0-Ip	Protruded, pedunculated
0-Is	Protruded, sessile
0-IIa	Superficial, elevated
0-IIb	Flat
0-IIc	Superficial, depressed
0-III	Excavated

Adapted from Endoscopic Classification Review Group 2005.

The limitations of conventional endoscopy and biopsy have led to the development of a number of advanced techniques for detection of dysplasia in Barrett's oesophagus, which will be discussed further in section 1.10 below.

1.7. Surveillance for Barrett's oesophagus

1.7.1. Guidelines

Patients with Barrett's oesophagus may be offered routine endoscopic surveillance with the aim of early detection of dysplastic or malignant change, to enable curative intervention. Any malignant change should ideally be detected prior to invasion of the submucosa, with its concomitant risk of lymphatic metastases.

To achieve this effectively requires accurate endoscopic and pathological recognition of such changes, along with a sufficiently short interval between endoscopies to monitor progression. These factors must be balanced against the acceptability and risks to the patient of multiple endoscopies, and the costs of implementing such a programme, particularly given the natural history of the condition which suggests a risk of progression to cancer below 0.5% per year (Bhat et al. 2011; Desai et al. 2012; Hvid-Jensen and Pedersen 2011).

Surveillance endoscopy programmes are now widely practised in many countries, however there remains uncertainty over the optimal surveillance strategy and this is reflected in the differing guidelines produced by a number of gastroenterological societies. A summary of the different societal guidelines for Barrett's surveillance is included in Appendix II.

The interval of surveillance endoscopy depends largely on the presence or absence of dysplastic change, but cases must be considered individually to account for patient preference and co-morbidity.

1.7.2. Efficacy of surveillance and cost-effectiveness

Evidence that endoscopic surveillance for Barrett's leads to improved outcomes (such as reduced mortality from oesophageal adenocarcinoma) is weak, and remains the subject of debate. The current evidence-base is derived from retrospective cohort studies and comparative studies: a number of small studies have shown improved outcomes and/or earlier stage at diagnosis in patients in surveillance programmes (Cooper et al. 2002; Cooper, Kou, and Chak 2009; Corley et al. 2002; Fountoulakis et al. 2004; Peters et al. 1994; Rubenstein et al. 2008; van Sandick et al. 1998; Streitz, Andrews, and Ellis 1993). However a recent, large retrospective study did not demonstrate any benefit from surveillance and reached the opposite conclusion (Corley et al. 2013). A comparison is shown in Table 1-6.

These retrospective studies are limited by a number of potential sources of bias and confounders, as the population presenting to medical professionals with symptoms or co-morbidities may be systematically different from those patients in whom Barrett's is not diagnosed and kept under surveillance. A randomised-controlled trial is needed to address this question, and a large RCT in the UK (Barrett's Oesophagus Surveillance Study) is underway to assess the efficacy and cost-effectiveness of surveillance.

Surveillance is a costly process, involving large numbers of patients having frequent endoscopies, and multiple tissue samples to be processed and reviewed by histopathologists. With hard data on efficacy lacking, there is a growing concern that surveillance is not cost-effective in its current forms (de Jonge et al. 2013; Kahrilas 2011; Vaezi and Kahrilas 2013).

Table 1-6 Comparison of retrospective studies of surveillance

Study	Total no. of patients with OAC	No. of patients detected by surveillance	Association with earlier cancer stage	Association with improved survival
Streitz et al. 1993	77	11	p=0.006	p=0.007
Peters et al. 1994	52	17	p=0.01	p=0.05
van Sandick et al. 1998	70	16	p=0.0001	p=0.0029
Corley et al. 2002	23	15	p=0.02	p=0.01
Cooper et al. 2002	1633	9.70%	p<0.001	p<0.01
Fountoulakis et al. 2004	91	17	p=0.001	p=0.008
Rubenstein et al. 2008	155	25	p=0.02	HR 0.82 (95% CI 0.52-1.29)
Cooper et al. 2009	2754	8.10%	p=0.001	p=0.001
Corley et al. 2013	351	70	(not stated)	HR 0.99 (95% CI 0.36-2.75)

Table adapted from Fitzgerald et al. 2014.

It must also be noted that, as a means by which to improve overall outcomes from oesophageal adenocarcinoma, surveillance can have only a limited impact, as a very small proportion of patients who develop adenocarcinoma have a prior diagnosis of Barrett's oesophagus. In a large, population-based study in Northern Ireland, only 7.3% of cases of oesophageal adenocarcinoma occurred in patients known to have Barrett's oesophagus (Bhat 2012).

Surveillance in its current form offers uncertain benefits and is not demonstrably cost-effective. The goal of current research is, firstly, to assess the benefits of surveillance and secondly, to enable risk stratification to identify those who may

benefit most from targeted surveillance at an appropriate interval. This knowledge should then allow the design of an efficacious and cost-effective programme of endoscopic surveillance for patients with Barrett's oesophagus.

1.8. Screening for Barrett's

The key criteria for a successful screening programme were outlined by Wilson and Jungner in a classic public health report in 1968 (Wilson and Jungner 1968), and have subsequently been updated in the UK by the National Screening Committee (National Screening Committee 2009).

Criteria for a successful screening programme (Wilson and Jungner 1968):

1 The condition sought should be an important health problem.

2 There should be an accepted treatment for patients with recognized disease.

3 Facilities for diagnosis and treatment should be available.

4 There should be a recognizable latent or early symptomatic stage.

5 There should be a suitable test or examination.

6 The test should be acceptable to the population.

7 The natural history of the condition, including development from latent to declared disease, should be adequately understood.

8 There should be an agreed policy on whom to treat as patients.

9 The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.

10 Case-finding should be a continuing process and not a “once and for all” project.

For oesophageal cancer, Barrett’s oesophagus represents the ‘recognizable latent or early symptomatic stage’ outlined above, and as such is a potential target for screening. However there are several criteria that are not met: the current gold-standard test is endoscopy and biopsy, which is neither ‘acceptable’ given that the small risk of morbidity would not be justified unless there was a sufficiently high prevalence amongst the screened population, nor a cost-effective means of case-finding.

These obstacles could be overcome in one of two ways: identification of ‘high-risk’ groups in which the associated morbidity and cost of endoscopy is justified, or through developing an alternative test with lower morbidity and lower cost.

To target those most at risk of having Barrett’s oesophagus several groups have examined the possibility of screening patients with chronic symptoms of gastro-oesophageal reflux disease (GORD) with endoscopy. The prevalence of Barrett’s in patients with chronic GORD is 5-15% (Shaheen and Ransohoff 2002). A meta-analysis by Taylor and Rubenstein (2010) found that, whilst there was a strongly increased risk of long-segment Barrett’s oesophagus with GORD symptoms (OR 4.92, 95% CI 2.01-12.0), there was no association between short-segment Barrett’s oesophagus and GORD (OR 1.15, 95% CI 0.76-1.73) (Taylor and Rubenstein 2010).

These symptoms are common amongst the general population however, with around 20% of the population having symptoms up to once a week (Shaheen and Ransohoff 2002), and around 6% of the population over the age of 45 experiencing chronic symptoms (Camilleri et al. 2005). To screen all of these

individuals would require huge resources: estimates from the US suggest that 6.6 million individuals would require endoscopy, with 1 oesophageal adenocarcinoma detected for every 1320 procedures (Kahrilas 2011; Vaezi and Kahrilas 2013). It must also be noted that 40% of patients with oesophageal adenocarcinoma (and 71% of patients with junctional adenocarcinoma of the cardia) do not experience any GORD symptoms (Lagergren et al. 1999).

Consequently endoscopic screening in an unselected population with GORD is not currently recommended either in the US or the UK (Fitzgerald et al. 2014; Spechler et al. 2011; Wang and Sampliner 2008). However, in the presence of multiple risk factors (chronic GORD plus 3 out of male sex, age over 50, white race, obesity) the BSG guidelines advise consideration of screening, with a lower threshold for those with a first-degree relative with Barrett's or oesophageal adenocarcinoma (Fitzgerald et al. 2014).

There has been recent focus on using genetic markers of risk to identify those at increased risk of developing OAC. A genome wide association study identified genomic markers of risk for Barrett's oesophagus, incorporating data from a number of studies of patients with Barrett's oesophagus (ChemoPrevention Of Intestinal Neoplasia (ChOPIN), Aspirin Esomeprazole Chemoprevention Trial (AspECT), and the Esophageal Adenocarcinoma Genetics LinkagE (EAGLE) consortium) (Su et al. 2012). Single nucleotide polymorphisms (SNPs) that increase the risk of BO were identified at chromosome loci 6p21, rs9257809, and 16q24, rs9936833. These SNPs could be used in future to identify those at increased risk of developing OAC and thus receive targeted screening.

Further work on genetic biomarkers of disease progression has identified markers of increased risk for progression to dysplasia or adenocarcinoma (p53 and SMAD4). This knowledge could be applied to risk stratify those with Barrett's from whom tissue or cell samples show these key mutations.

An alternative approach to screening is to use less invasive techniques to reduce the morbidity. Studies assessing capsule endoscopy to identify Barrett's have shown a relatively low sensitivity at 60-78% (Galmiche et al. 2008; Lin et al. 2007; Ramirez, Akins, and Shaukat 2008). Others have used ultra-thin transnasal endoscopes in patients with GORD or known Barrett's and demonstrated results very similar to conventional endoscopy (Jobe et al. 2006; Shariff et al. 2012).

A further technology under assessment is a swallowed cytology collection device (Cytosponge™) (Kadri et al. 2010). The device is contained within a small capsule which is attached to a length of string: patients swallow the capsule, the gelatine capsule then dissolves in the acidic gastric secretions to reveal the collection device (analogous to a cytology brush), which is then withdrawn through the oesophagus using the string, collecting cells as it passes (see Figure 1-15). Cytology collection has the potential advantage over biopsy of sampling very large areas of epithelium, to minimise the risk of missing small areas of dysplasia.

Accurate biomarkers of Barrett's oesophagus are required due to the mixed cell population collected by the device, with a previous study suggesting the most promising as Trefoil Factor 3 (TFF3) (Lao-Sirieix et al. 2009). Furthermore, the acceptability of the device has been demonstrated in a primary care setting (Kadri et al. 2010).

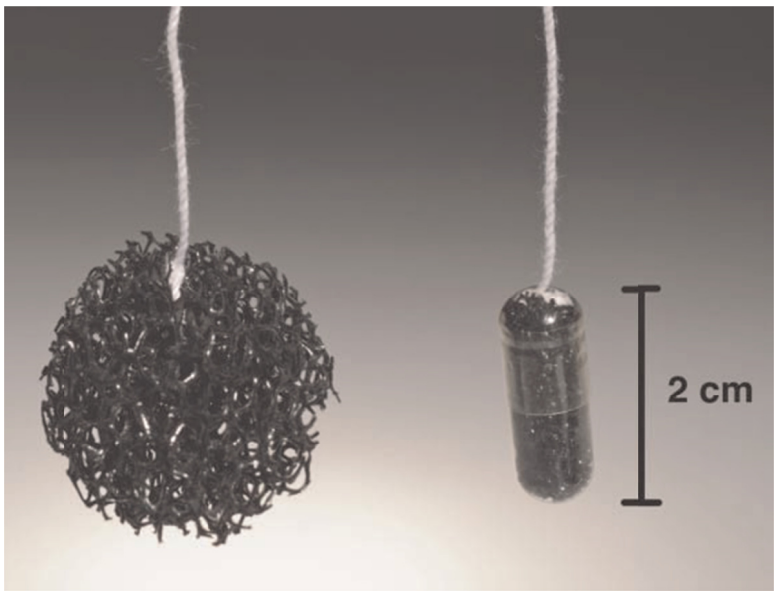


Figure 1-15 Cytosponge™: a swallowed cell collection device. Reproduced from (Kadri et al. 2010).

At the time of our study design, a trial was underway to evaluate the Cytosponge™ in a case-control study in the UK (Barrett's Esophagus Screening Trial 2, BEST2). This trial has recently reported and these results are discussed in the Summative Discussion (Chapter 10).

To date, none of these technologies have entered routine clinical practice, but screening could radically alter both the management approach to Barrett's and the burden of disease to be managed.

1.9. Management of Barrett's and Barrett's-associated neoplasia

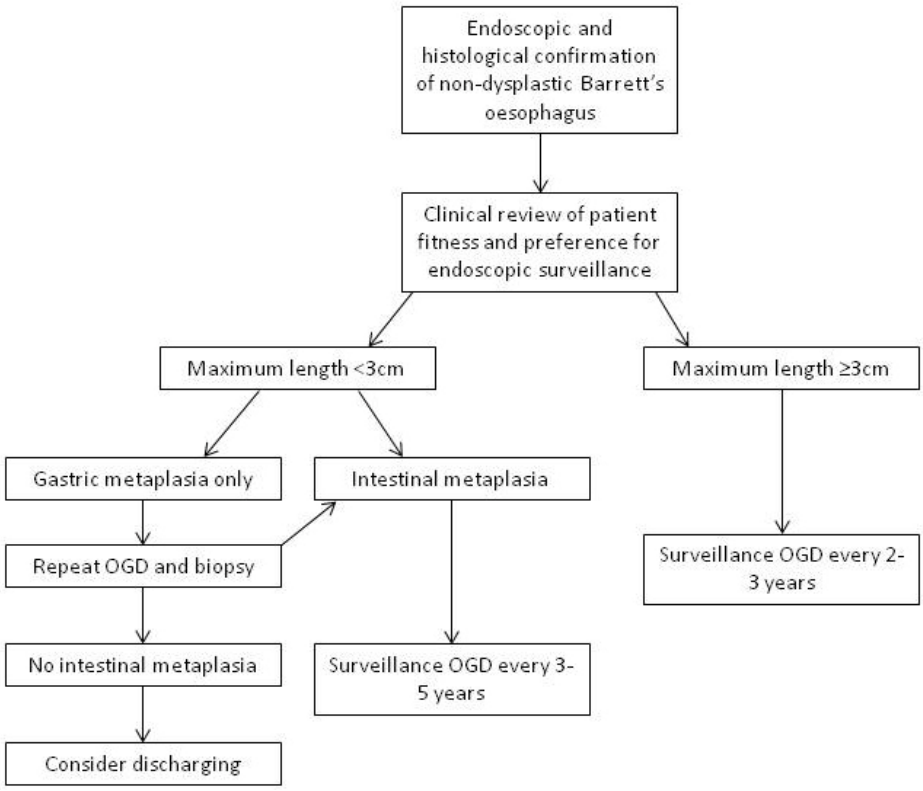


Figure 1-16 Management of non-dysplastic Barrett's oesophagus (adapted from Fitzgerald et al. 2014).

Surveillance forms the mainstay of management for patients with non-dysplastic Barrett's oesophagus: the current BSG guidelines are illustrated in the algorithm in Figure 1-16 above. Endotherapy is not currently indicated for these patients, but there is emerging evidence to support a role for chemoprevention, outlined below.

1.9.1. Chemoprevention

There is increasing evidence for the use of certain drugs as chemoprevention for patients with Barrett's oesophagus. A number of cohort studies have shown a significantly reduced risk of progression to high grade dysplasia (HGD) or

oesophageal adenocarcinoma (OAC) for patients taking proton pump inhibitors (PPIs) versus no therapy or on histamine-2 receptor antagonists (H2RA) (El-Serag et al. 2004; Nguyen et al. 2009). However data from randomised controlled trials (RCT) are still awaited, and the BSG guidelines advise that 'there is not yet sufficient evidence to advocate acid suppression drugs as chemopreventive agents' (Fitzgerald et al. 2014).

Evidence for a protective effect from non-steroidal anti-inflammatory drugs (NSAIDs) comes from large meta-analyses of patients taking aspirin as primary or secondary prevention for cardiovascular disease (Corley et al. 2003; Liao et al. 2012; Rothwell et al. 2011). In the largest of these, incorporating data from 23,535 patients, those followed up for 10-20 years after starting aspirin, and having taken aspirin daily for 5 years or more, had a significant reduction in risk of OAC, hazard ratio 0.36 (95% CI 0.18-0.71) (Rothwell et al. 2011).

Once again there are no data from randomised trials, and given the risks of NSAIDs such as gastrointestinal bleeding and cerebral haemorrhage, this will be crucial in informing management decisions for patients with Barrett's (Cuzick et al. 2009).

Aspirin and Esomeprazole for Chemoprevention in Barrett's metaplasia (AspECT) is a large, multi-centre RCT (NCT00357682), which aims to address the lack of randomised data and evaluate the risks and benefits of both PPIs and NSAIDs as chemoprevention in Barrett's (ClinicalTrials.gov 2013). The trial has 4 arms, with patients randomised to low-dose or high-dose PPI (esomeprazole) +/- aspirin.

There is some evidence from observational studies for a protective effect from statins against oesophageal cancer. A recent meta-analysis reviewed the

effects of statins in a patients with Barrett's (Beales, Hensley, and Loke 2013), including 11 observational studies with a total of 1999 patients, and found an OR for progressing to OAC was 0.57 (95%CI 0.43-0.75) for those on a statin.

Whilst the absence of randomised data limits current recommendations for chemoprevention, this paradigm may become a key component of management of Barrett's oesophagus in future.

1.9.2. Endotherapy for Barrett's-associated neoplasia

Endoscopic treatments aim to remove or destroy areas of neoplasia in the oesophagus, either through resection or ablation, and promote re-growth of the normal squamous lining. The goal of ablative techniques is to achieve complete eradication of dysplasia, along with complete eradication of intestinal metaplasia (Leggett, Gorospe, and Wang 2013). These minimally invasive treatments can offer treatment without recourse to radical surgery and thus can be considered in patients with earlier stage disease and those for whom major surgery would be a very high risk.

The BSG guidelines advise the following algorithm (see Figure 1-17) for management of dysplastic Barrett's, and this is discussed further below.

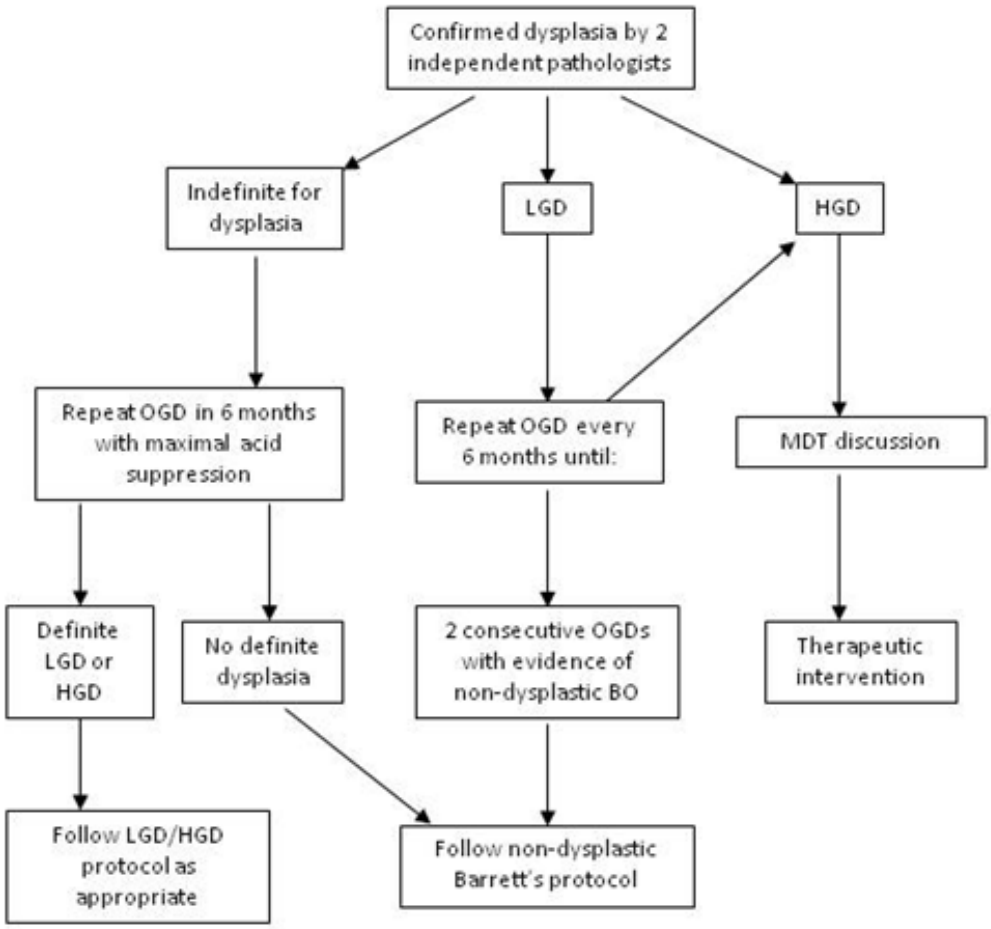


Figure 1-17 Algorithm for management of dysplastic Barrett's oesophagus (adapted from Fitzgerald et al. 2014). LGD low grade dysplasia; HGD high grade dysplasia; OGD oesophago-gastroduodenoscopy; MDT multidisciplinary team

1.9.2.1. Low grade dysplasia

In the 2014 BSG guidelines, as shown in figure 1-17 above, patients found to have LGD are advised to have 6-monthly OGD surveillance until biopsy confirms regression to non-dysplastic BO on successive OGDs, or progression to HGD occurs (Fitzgerald et al. 2014). However, a recent randomised controlled trial from European centres reported a highly significant reduction in

progression for patients with LGD treated with RFA (Phoa et al. 2014). The rate of progression to HGD or cancer over 3 year follow up was 1.5% in the ablation group versus 26.5% in the untreated group ($p<0.001$), with progression to cancer 1.5% versus 8.8% respectively ($p=0.03$).

It must be noted that the rate of progression in the control arm in this trial (26.5%) was high over 3 years, which may limit its applicability in other settings. This rate is much higher than might be expected from the natural history of LGD reported in other studies, with a recent meta-analysis calculating a risk of progression to HGD or cancer of around 1% per year (Almond, Hodson, and Barr 2014). The use of an expert pathology panel for consensus diagnosis, with a very robust classification of LGD, is likely to have influenced the composition of the group seen in the study by Phoa et al. Of 511 patients with an initial diagnosis of LGD screened for entry to the trial, only 247 were confirmed to have LGD after review by the expert panel, and 140 went on to be randomised (Phoa et al. 2014).

In light of these findings it seems likely that historical overdiagnosis of LGD has led to an underestimation of the true risk of progression with LGD (Almond et al. 2014). This suggests the need for consensus reporting of all Barrett's dysplasia specimens, and the question of whether to offer intervention in the absence of consensus is difficult.

The recently published National Institute for Health and Care Excellence (NICE) guidelines now support the use of RFA for the ablation of low grade dysplasia (National Institute for Health and Care Excellence 2014).

1.9.2.2. High grade dysplasia

The diagnosis of HGD has serious implications for patients – not only do they have a high risk of progression to adenocarcinoma, but there is a significant risk that a small focus of cancer may already be present, but not yet detected. For this reason, historically, the treatment of HGD was radical oesophagectomy for those considered fit for surgery. The advent of endotherapy led to great debate over whether these treatments could offer oncologically sound treatment. Although there have not been any randomised trials directly comparing surgery with endotherapies, the growing body of evidence for endotherapies over the past 20 years or so has resolved this debate, and endoscopic resection is now considered the treatment of choice for patients with macroscopically visible HGD or T1a adenocarcinoma (Fitzgerald et al. 2014). The management of HGD and early oesophageal cancer is shown in the algorithm in Figure 1-18 below.

For visible lesions, endoscopic resection (ER) is preferred over ablative therapy alone because it provides the most accurate staging information. Depth of invasion can be accurately assessed: data from stepwise ER of entire Barrett's segments have confirmed that the most advanced disease is located in visible lesions (Pouw et al. 2010). A meta-analysis of studies in which patients have undergone oesophagectomy for HGD found the risk of invasive OAC to be 11% in those with visible lesions, compared to 3% in those with no visible lesion (Konda et al. 2008).

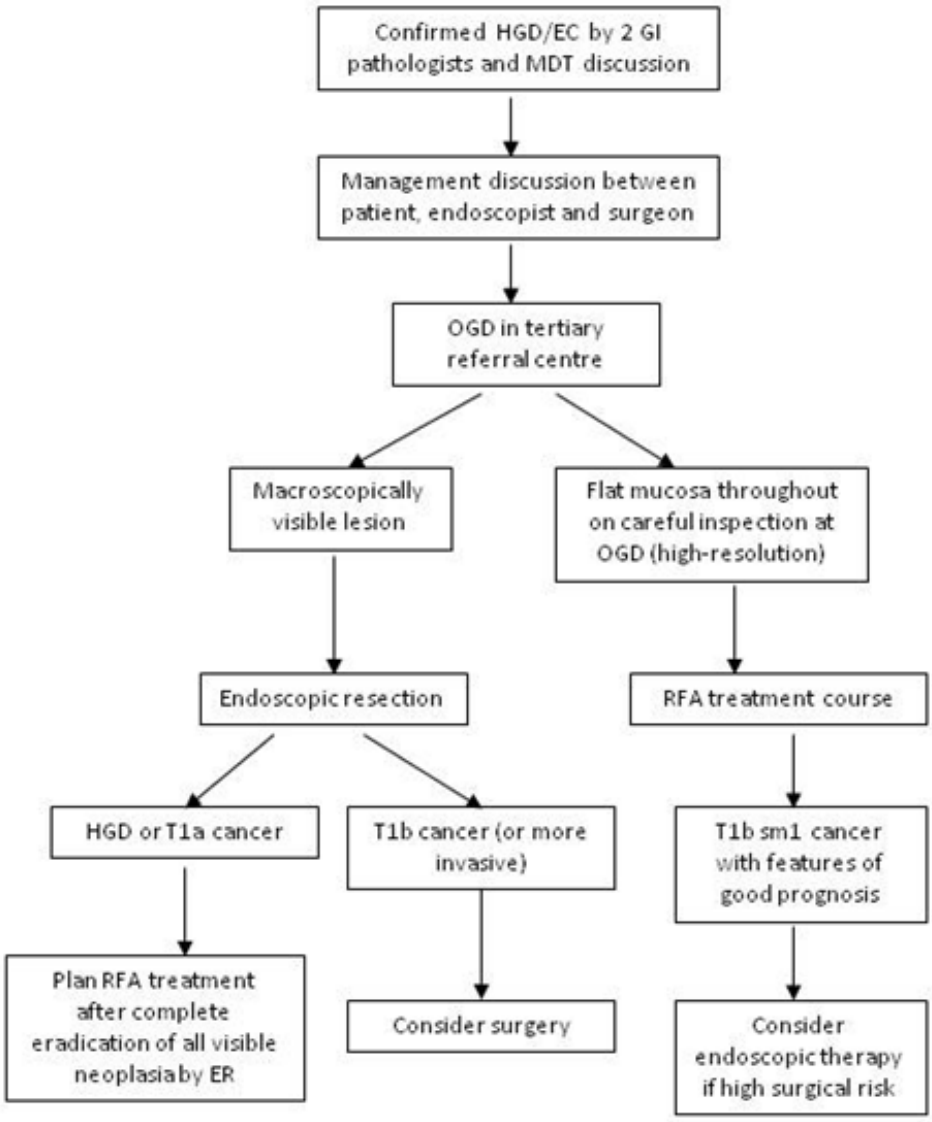


Figure 1-18 Algorithm for managing HGD and early oesophageal adenocarcinoma (adapted from Fitzgerald et al. 2014). ER endoscopic resection; HGD high grade dysplasia; EC early oesophageal adenocarcinoma; MDT multidisciplinary team; OGD oesophagogastroduodenoscopy; RFA radiofrequency ablation

There are now numerous ablative techniques for HGD: each of these techniques has its strengths and weaknesses, and some key features are presented in Table 1-7 below. RCT data are available for some of these modalities, and includes direct comparison of certain techniques. Currently,

there is a large scale trial (Barrett's Radiofrequency Intervention for Dysplasia by Endoscopy, BRIDE, NCT017337) under way aiming to compare outcomes between RFA and argon plasma coagulation (APC) (ClinicalTrials.gov 2012). This pilot study has just closed to recruitment and initial data will be available soon. On the basis of currently available RCT data and taking account of the side effect profiles of each treatment, RFA is recommended as the first-line therapy for HGD outside of the context of RCTs (Fitzgerald et al. 2014).

The first major RCT of RFA randomised 127 patients on a 2:1 allocation to RFA or a sham procedure (Shaheen et al. 2009). Complete eradication of HGD was seen in 81.0% of treated patients versus 19.0% of controls ($p<0.001$). In a later report on extended follow up of this group, 96% of those achieving eradication of HGD remained free of HGD at 3 year follow up (Shaheen et al. 2011). Risk of any disease progression was reduced in the ablation group (3.6% vs. 16.3%, $p=0.03$) as was the risk of cancer (1.2% vs. 9.3%, $p=0.045$).

A meta-analysis of RFA for BO calculated pooled estimates for complete eradication of dysplasia of 91% (87-95%, 95% CI), and complete eradication of IM 78% (70-86%, 95% CI) (Orman, Li, and Shaheen 2013). The commonest complication was stricture, which occurred in 5% of patients treated with RFA.

Table 1-7 Comparison of ablative techniques for HGD

Technique	Eradication of dysplasia	Advantages	Disadvantages
Step-wise radical ER	97-100%	Accurate diagnosis Low costs	High risk of stenosis Only feasible <5cm Barrett's
APC	67-86%	Widely available Low costs	Buried glands Feasible for short segments only
PDT	40-77%	RCT data Treatment of nodular dysplasia	High risk of stricture Buried glands Photosensitivity
RFA	80-98%	RCT data High response rate Low complication rate	High costs Minimal long-term follow up data
Cryotherapy	68-88%	Good safety profile	No RCT data No long-term follow up data

Adapted from Fitzgerald et al., 2014. APC, argon plasma coagulation; ER endoscopic resection; PDT photodynamic therapy; RCT Randomised controlled trial; RFA radiofrequency ablation.

1.9.2.3. Intra-mucosal adenocarcinoma

The low risk of nodal metastasis with T1a cancer (0-10%) (Alvarez Herrero et al. 2010; Barbour et al. 2010; Buskens et al. 2004; Liu et al. 2005; Prasad et al. 2009; Sepesi et al. 2010; Stein et al. 2005; Westerterp et al. 2005) has resulted in very good tumour free and overall survival with endoscopic resection for T1a cancers in high-volume centres. One recent large series reported tumour free

survival of 93.8% at mean follow up 56.6 months (Pech et al. 2014). ER for T1b tumours is associated with poorer outcomes, with 5 year tumour free and overall survival 60% and 58% respectively, likely due to increased depth and higher risk of nodal involvement (up to 46%) (Alvarez Herrero et al. 2010; Barbour et al. 2010; Buskens et al. 2004; Liu et al. 2005; Prasad et al. 2009; Sepesi et al. 2010; Stein et al. 2005; Westerterp et al. 2005). For upper third tumours of the submucosa (T1bsm1) the risk of lymph node metastasis is relatively low (around 10%) and some series report good outcomes treating T1bsm1 tumours endoscopically (Alvarez Herrero et al. 2010; Pech et al. 2014; Westerterp et al. 2005). Decision making may be further influenced by pathological indicators of good prognosis, such as R0 resection, or absence of vascular and lymphatic invasion (Estrella et al. 2011). The use of ER for T1bsm1 cancer remains debated however, and the current BSG guidance advises surgery for patients who are fit enough, but that ER should be offered with curative intent in patients who are high-risk surgical candidates (see Figure 1-18 above) (Fitzgerald et al. 2014).

Whilst ER is effective at removing Barrett's-associated intra-mucosal lesions, these patients have a high rate of metachronous lesions after ER alone, affecting around 15-20% (Pech et al. 2008, 2014). The risk of developing further lesions can be reduced significantly with ablative therapy to the remaining Barrett's segment (Pech et al. 2008; Pouw et al. 2010). Current guidance advises removal of all visible lesions with ER, followed by ablative therapy (currently RFA is preferred method) to all residual areas of Barrett's (Fitzgerald et al. 2014).

1.10. Advanced endoscopic diagnosis

1.10.1. Wide-Field detection techniques ('red-flag')

A number of techniques have been developed to increase the sensitivity of detecting areas of dysplasia or neoplasia during wide-field endoscopic examination, as opposed to those techniques which aid histological recognition on a point-by-point basis, analysing small areas of tissue at a time.

1.10.1.1. High Resolution Endoscopy

The increasing resolution of cameras (>1,000,000 pixels) and high definition display screens has improved detection over conventional equipment. Current systems allow magnification up to x150 depending on equipment (Subramanian and Ragunath 2013). High resolution endoscopes (HRE) and monitors increase sensitivity, with some studies reporting HRE to be as good as NBI and chromoendoscopy, with no additional benefit from the enhanced imaging techniques (W. Curvers et al. 2008). It is also widely available in clinical use. However there are a number of weaknesses compared to alternative advanced imaging techniques: sensitivity for dysplasia can be variable (79-90%) and shows interobserver variability (W. Curvers et al. 2008; Kara et al. 2005), plus longer inspection times are required to achieve these higher detection rates (Gupta et al. 2012).

1.10.1.2. Chromoendoscopy

Applying a 'stain' (most commonly acetic acid or methylene blue) via the endoscope to the oesophageal mucosa enables enhanced visualisation,

highlights surface topography and mucosal irregularities. Acetic acid particularly accentuates villi and pits and can increase detection of intestinal metaplasia and dysplasia compared to white light endoscopy (WLE) (Fortun et al. 2006). Methylene blue highlights intestinal metaplasia and dysplasia but not squamous epithelium. Reported sensitivity for identifying IM ranges from 53-98%, with specificity 32-97% (Saxena and Canto 2013). This wide range is likely to be due to study heterogeneity, different administration techniques and methods of interpretation when performing chromoendoscopy. The low specificity, interuser variability, and practicalities of administration of stains (additional catheter and messy dye) have limited the adoption of this technique, particularly as it has largely been overtaken by virtual chromoendoscopy.

1.10.1.3. Narrow Band Imaging

The addition of an extra light filter to endoscopic white light narrows the bandwith of blue and green light, and increases the intensity of the blue light (Saxena and Canto 2013). The blue light does not penetrate mucosa, and the blue wavelength corresponds to the peak absorption of haemoglobin, therefore highly vascular structures appear darker, contrasting with the surrounding mucosa (Subramanian and Ragunath 2013). Some studies have shown high accuracy in experienced hands, with a meta-analysis of 8 studies showing high sensitivity (96%) and specificity (94%) for HGD, high sensitivity (0.95) but poor specificity (0.65) for IM (Mannath et al. 2010). There is also evidence for equivalent detection rate to WLE but achieved with fewer biopsies (Wolfsen et al. 2008). However there are wide-ranging figures quoted on real-time sensitivity and specificity depending on methodology used: one multi-centre

study with expert endoscopists showed sensitivity for HGD 83% and specificity 33% (Curvers, Herrero, et al. 2010). A further multi-centre study reported sensitivity for dysplasia 47% and specificity 72% (Singh et al. 2011), and noted that interobserver agreement was better amongst experts (k=0.63 for mucosal changes, k=0.58 for vascularity and k=0.68 for predicted histology) than non-experts. One of the causes for interobserver variability is the use of multiple classification systems, all of which have their own limitations (Silva et al. 2011; Singh et al. 2011). It is also relatively time-consuming compared to standard endoscopy (Curvers, Herrero, et al. 2010).

1.10.1.4. FICE (Fuji Intelligent Chromo Endoscopy)/i-scan (Pentax)

This system is analogous to NBI but use post-processing techniques to enhance the image rather than light filters to improve visualisation of mucosal and vascular patterns. The strengths are likely to be similar to NBI, but whilst some work has been done in the colon, there have been few studies examining its role in Barrett's to date. One study showed FICE and acetic acid chromoendoscopy identified irregular mucosal patterns in 100% of patients in a small pilot study (Camus et al. 2012).

1.10.1.5. Autofluorescence

Autofluorescence imaging (AFI) produces real-time pseudocolour images based on detection of fluorescence from endogenous fluorophores when exposed to short wavelength light (370-470nm) (Saxena and Canto 2013; Subramanian and Ragunath 2013). Neoplastic tissues have different autofluorescence

characteristics, allowing their differentiation: a barrier filter shows normal tissue as green and dysplastic/cancerous tissue as magenta. It appears to be highly sensitive, with several authors reporting sensitivity >90% (Curvers et al. 2008a; Kara et al. 2006). The major weakness has been poor specificity, with a high number of false positives (difficulty discriminating dysplasia and inflammation) (Curvers et al. 2008a). Attempts have been made to improve imaging techniques but poor specificity remains a problem for 'third-generation' as well as 'second-generation' AFI (Boerwinkel et al. 2013). This combination of characteristics lends itself to use with other imaging modalities as a 'red-flag' technique, sampling large fields which can then be analysed further with more specific techniques.

Future developments may include multimodal systems, attempts to address the low-resolution image and jerky video because of low frame rate of current systems (Subramanian and Ragunath 2013), or attempts to correlate AFI with biomarkers, though this has not been very successful to date (Boerwinkel et al. 2012).

1.10.1.6. Labelling of biomarkers

The concept of labelled biomarkers has excited a lot of interest: the aim is to use visually tagged probe molecules that selectively bind neoplastic cells. If reliable biomarkers could be identified, there are a number of potential techniques that could be used to visualise the markers and thus highlight areas of neoplasia.

Probe molecules that would target such biomarkers may either be antibodies against specific proteins on neoplastic cells, or less specific molecules that selectively bind to, or concentrate in, neoplastic cells. Probe molecules could be labelled in a number of ways, including fluorescence, or binding to nanoparticles which carry a label e.g. Surface Enhanced Raman Scattering (SERS) using a metallic nanoparticle with a chemiadsorbed Raman reporter molecule (Kong et al. 2015).

One example of fluorescence in Barrett's oesophagus is lectins – alterations in cell surface glycans cause changes in lectin binding to selectively target neoplasia, with the use of fluorescent lectins allowing visualisation (Bird-lieberman et al. 2012).

These techniques have the potential to rapidly identify areas of dysplasia and neoplasia, with many putative targets. However, to date, although there has been some *in vivo* work to demonstrate proof of concept, sensitive and specific markers are lacking (Lu and Wang 2008).

1.10.1.7. Multimodal imaging

To utilise the different advantages of differing imaging modalities, there has been much work on multimodal imaging systems, whereby a combination of techniques can be used concurrently to aid diagnosis.

Whilst a number of combinations have been tried, the most successful to date has been the Endoscopic TriModal Imaging system (Olympus). This incorporates high-resolution endoscopy, AFI and NBI. The aim is to use AFI initially, which has a high sensitivity for dysplasia, to identify suspicious lesions,

followed by NBI which has greater specificity to characterise them further and reduce false positives (the main problem associated with AFI alone) (Curvers, Herrero, et al. 2010; W. L. Curvers et al. 2008). The theory follows that AFI should avoid missing any suspicious lesions, and NBI should prevent a high number of unnecessary biopsies. However, a recent randomised controlled trial in the Netherlands found that although there was a greater yield of targeted biopsies, NBI appeared to add little, and the overall diagnostic yield was not significantly different from conventional endoscopy (Curvers, Herrero, et al. 2010).

Certain other weaknesses also remain – it does not define the histology and therefore still requires biopsy confirmation, use of multiple modalities is time consuming, and there is the increased cost of a multimodal system.

Another possible multimodal technique combined cross-polarized imaging (CPI), vital dye fluorescence imaging (VFI) and high resolution microendoscopy (HRME) (Thekkek et al. 2013). CPI shows vascularity (like NBI), VFI shows glandular structures. After these wide-field techniques they applied point measurement with HRME. This demonstrated *in vivo* feasibility, but there are no published data on sensitivity and specificity.

1.10.2. Point measurement techniques

1.10.2.1. Confocal laser endomicroscopy

Confocal laser endomicroscopy (CLE) magnifies mucosa x1000, making it possible to visualise cellular structures in real-time. The microscope may be integrated into the endoscope (Pentax systems) or incorporated into a probe

which is passed via the instrument channel (Cellvizio™). In experienced hands the technique can have accuracy as high as 96.8% for IM and 92.9% for intraepithelial neoplasia in the oesophagus (Goetz and Kiesslich 2008). It defines tissue histology, and consequently may allow a reduction in the number of biopsies required (Canto et al. 2013). In one recent study, compared to WLE alone, CLE altered the patient's management plan in 36% of cases (Canto et al. 2013). Weaknesses of this technique include interuser variability, subjective interpretation, and it is time-consuming as small areas are sampled. It is also comparatively expensive and requires intravenous fluorescein.

The similar technique of endocytoscopy uses a contact light microscope to give up to x1400 magnification. Recent *ex vivo* work in BO has shown high accuracy (83-100%) and high interobserver agreement for this technique (Tomizawa et al. 2013).

1.10.2.2. Optical Coherence Tomography

Optical coherence tomography (OCT) is analogous to ultrasound using electromagnetic waves: it produces its image from infrared light and optical scattering based on differences in tissue composition (Lightdale 2013). Techniques used have included a balloon catheter system (Suter et al. 2009) or a tethered capsule (Gora et al. 2013). In addition to defining the histology, OCT provides cross-sectional imaging, which could potentially be useful for assessing depth of invasion of a lesion, or detecting buried glands following ablative treatment in the oesophagus (Lightdale 2013). However it has poor image contrast, comparatively low sensitivity and specificity (one study found the sensitivity for HGD/IMC to be 83%, specificity 75% (Evans et al. 2006), and

a further study by the same group reported sensitivity 81% and specificity 57-66% for IM) (Evans et al. 2007). This may in part be due to the difficulties and somewhat subjective interpretation required.

1.10.2.3. Elastic Scattering Spectroscopy (ESS)

This technique is based on the elastic scattering of white light: light scattering from epithelial surfaces provides information about cellular organelles e.g. nuclear size. Development of a fibre-optic probe has been the aim, with advantages being a short acquisition time, relatively low cost, and an intermediate field of measurement at around 0.5mm. High signal interference from deep structures has proved a limitation on accuracy, but the technique has shown sensitivity for HGD (92%) though not specificity (60%) (Lovat et al. 2006).

1.10.2.4. Endoscopic polarised scanning spectroscopy (EPSS)

EPSS is similar to ESS above, depending on elastically scattered light, but this technique uses a filter to remove signal from deeper tissue to allow epithelial analysis (light backscattered from the epithelial surface retains its linear polarisation, unlike light backscattered from the submucosal layers) (Qiu et al. 2010). This may achieve higher accuracy than ESS: Qiu et al. report sensitivity of 92% and specificity of 96% for all pathologies in a study of 8 patients *in vivo,* but current data are limited.

1.10.2.5. Angle-resolved low coherence interferometry

This is a further optical technique based on angular scattering from cell nuclei at various tissue depths, which provides a measure of the size of cell nuclei, with larger nuclei implying dysplasia. A small *in vivo* feasibility study has demonstrated some promise for identification of dysplasia (Wax et al. 2011).

A summary table comparing advanced endoscopic imaging techniques is included in Appendix III.

CHAPTER 2: VIBRATIONAL SPECTROSCOPY

2. Vibrational spectroscopy

2.1. Fundamental principles

When light is incident on matter it may be reflected, absorbed (including fluorescence and phosphorescence), scattered, or pass straight through. Incident light is absorbed by a molecule if its energy matches the energy required to excite the molecule from its ground state to an excited state. Only light of a specific wavelength (and thus a specific energy) will be absorbed by a given molecule. Scattering occurs when the energy from incident light raises the electrons of the molecule to a higher energy state, forming an unstable 'complex', before releasing this energy as scattered radiation. Nearly all of this scattered radiation has the same wavelength (and hence energy) as the incident light: since the incident and scattered light has the same energy this is termed 'elastic' scattering, or known as Rayleigh scattering. Around 1 in 10^6-10^8 of the scattered photons have a different energy from the incident light, corresponding to one of the vibrational energy states of the molecule: this is termed 'inelastic' scattering, or known as Raman scattering. When the molecule starts in its ground state, and the scattered light has a lower energy than the incident light, the process is known as Stokes scattering. If the molecule begins at an excited state and the scattered light has a greater energy than the incident light, it is termed Anti-Stokes scattering. The energy changes involved in these processes are summarised in Figure 2-1 below.

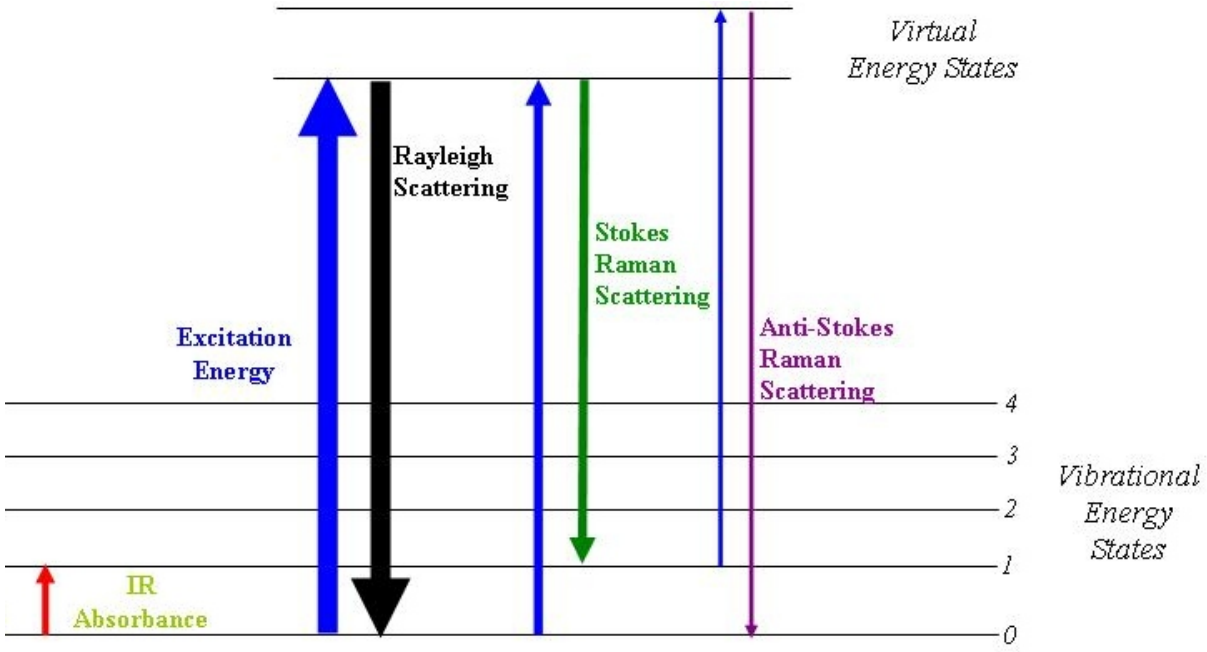


Figure 2-1 Changes in energy state underlying infrared absorbance, Rayleigh and Raman scattering. (reproduced from http://commons.wikimedia.org/wiki/File:Raman_energy_ levels.jpg)

Changes in energy state will alter molecular vibration. A given molecule can vibrate in a defined number of vibrational modes: for a non-linear molecule this is 3N-6, where N is the number of atoms in the molecule (see Figure 2-2 below). For infrared absorption to occur, molecular vibration must induce a change in the dipole moment of the molecule. For Raman scattering to occur, there must be polarisation of the electron cloud. Different molecules require different energy to alter these vibrational modes: this underpins the analytical technique of vibrational spectroscopy, with measurement of the energy change from infrared absorption or Raman scattering providing information about the molecules under analysis.

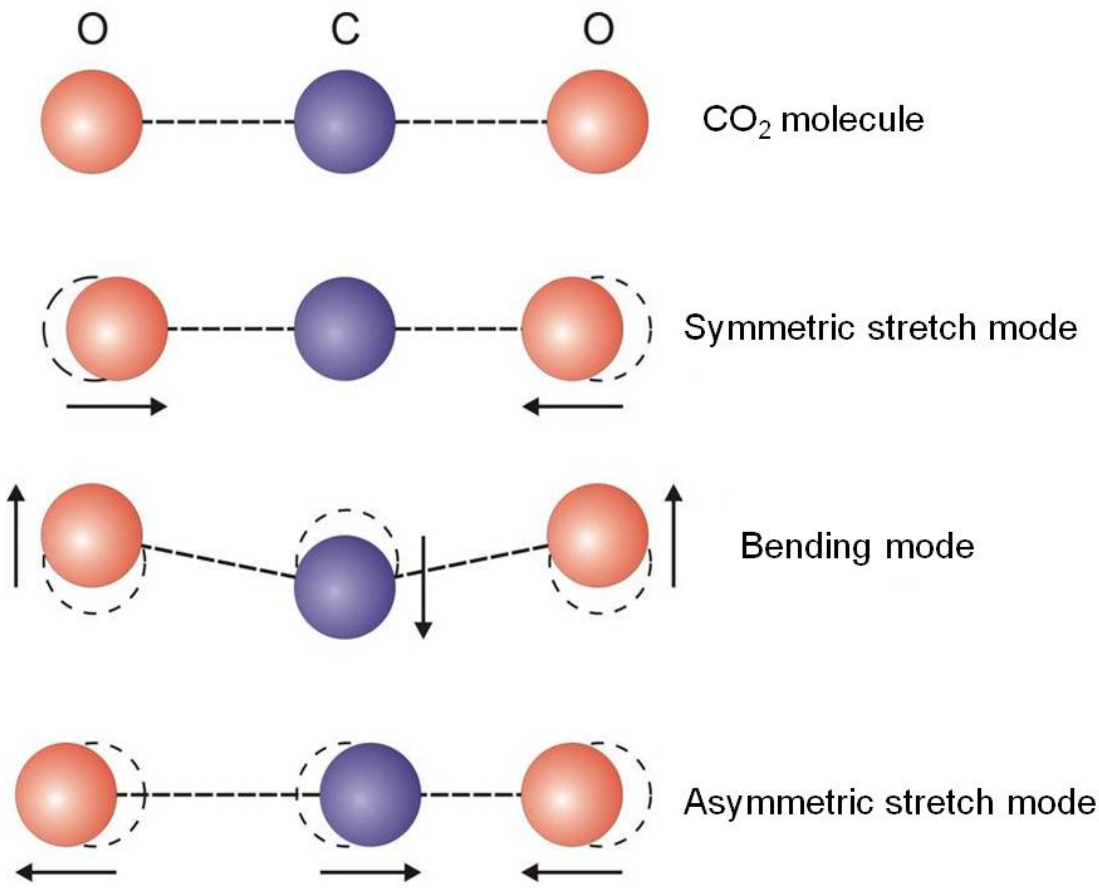


Figure 2-2 Possible vibrational modes for a CO_2 molecule. For a triatomic molecule such as CO_2, symmetric stretch has a strong effect on the polarisation of the electron cloud, and hence a relatively strong Raman scattering effect. Bending and asymmetric stretch cause a change in the dipole moment and consequently a strong infrared absorption, but a weak effect on polarisation and correspondingly weak Raman scattering. (adapted from http://www.ino.it/~miomao/co2modi.jpg)

2.2. Raman Spectroscopy

2.2.1. Raman spectroscopy of biological samples

Raman spectroscopy relies on the inelastic (Raman) scattering of light to interrogate samples and analyse their content. The difference in energy

between incident and scattered photons is known as the 'Raman shift'. This 'Raman shift' is both specific to the bond responsible for the scattering, and independent of the incident wavelength (Byrne, Sockalingum, and Stone 2011). The intensity of the scatter gives a measure of the concentration of the molecules containing responsible bonds in the specimen.

Plotting the Raman shift measured in wavenumbers (1/wavelength, conventionally in units of cm^{-1}) on the x-axis, against the intensity of the scattered light on the y-axis, gives the Raman spectrum for a given sample. Since the Raman shift is relative to the incident wavelength, the zero point on the x-axis represents the incident light, with positive values recording Stokes scattering, and negative values Anti-Stokes.

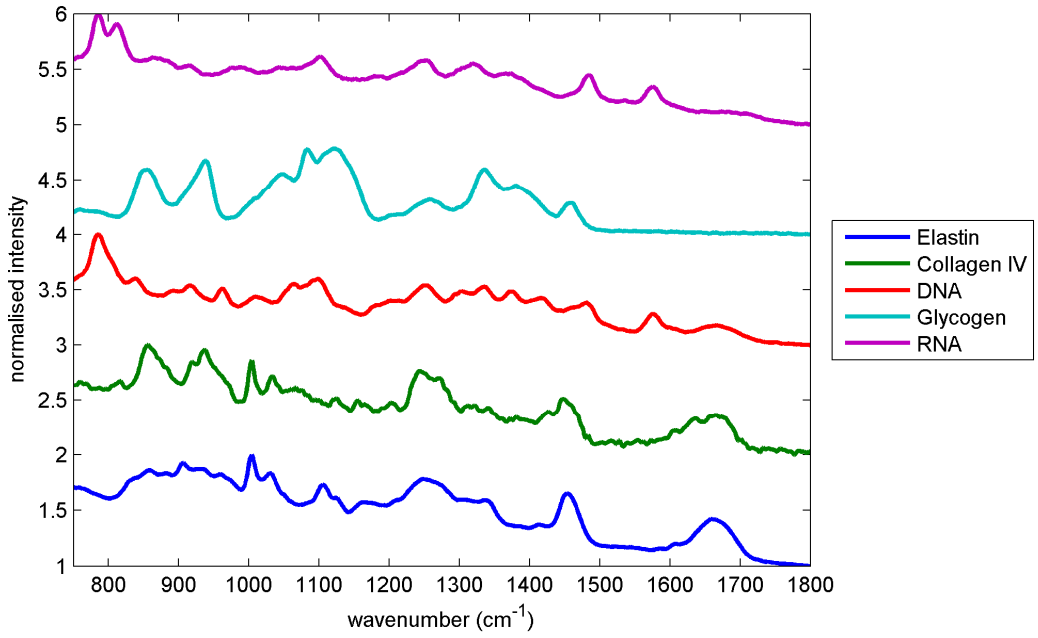


Figure 2-3 Raman spectra of certain key biomolecules – characteristic spectral peaks associated with specific biochemical substrates allow identification of molecules based on their spectra. Reproduced from Old et al. 2014.

The resultant spectrum is therefore both a qualitative and quantitative measure of the sample biochemistry (Byrne et al. 2011). Figure 2-3 shows an example of the use of Raman spectroscopy to analyse a number of key biomolecules.

Raman spectroscopy of biological samples poses a number of challenges: the signal is relatively weak, as only 1 in 10^6-10^9 photons are inelastically scattered and ambient light must be filtered to reduce unwanted spectral contributions (Kendall et al. 2009). Visible wavelengths excite a strong autofluorescence signal which can mask the weak Raman signal: near-infrared wavelengths are used to minimise this effect (Krafft 2009). Raman spectrometers must therefore be highly sensitive to Raman signal, whilst excluding all other signal through a series of filters. A schematic diagram of the layout of a typical laboratory Raman spectrometer is shown in Figure 2-4.

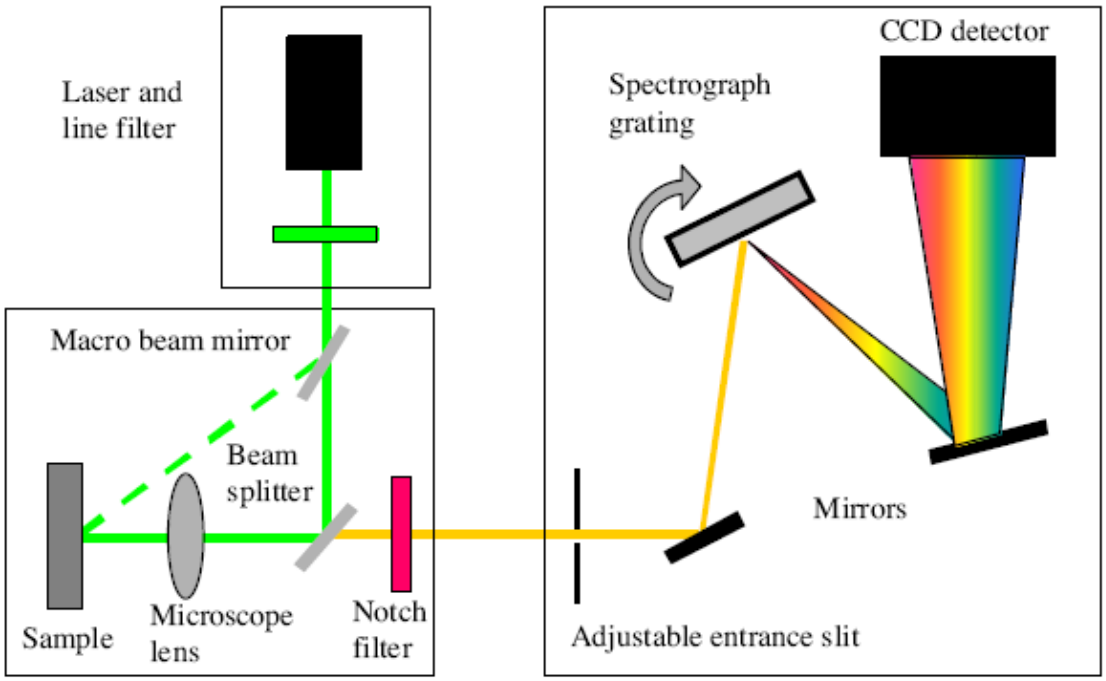


Figure 2-4 Layout of a Raman spectrometer (reproduced from http://www.sas.upenn.edu/~crulli/TheRamanSpectrophotometer.html)

The laser light is focussed onto the sample, then scattered light is filtered based on wavelength to remove any elastically scattered light, passes through a slit to remove any stray light, before a diffraction grating splits the beam based on its wavelength, and the light is then reflected onto a charge-coupled detector (CCD) device. The CCD registers the number of photons striking it at each region, which following diffraction corresponds to a specific wavelength, thus giving the intensity at each wavelength.

The -OH vibrational modes of water molecules produce weak Raman signals, and thus do not contribute significantly to resultant spectra. Consequently, Raman spectroscopy can be used to analyse fresh tissue with no prior preparation, both *ex vivo* and *in vivo* without causing tissue damage.

The use of fibre optic probes can improve tissue access for *in vivo* Raman analysis, such as through a needle probe or endoscope (Kast et al. 2007). Fibre-optic probes designed for *in vivo* use must overcome the signal-to-noise challenges, and face the major problem that Raman signals and photoluminescence are generated in the delivery fibres themselves (Day et al. 2009). For certain applications, all this must be achieved with a probe of very small size to enable access to body cavities, and in a sufficiently short time to allow accurate measurement from a moving target organ. This will be discussed in further detail with regard to the Raman probe used in this project.

Spectral analysis of biological tissue samples using vibrational spectroscopy has a number of potential strengths as a diagnostic tool. Data analysis techniques applied to the measured spectra can characterise subtle differences and detect biochemical changes that have been induced by disease. This enables identification of pathology, and classification into appropriate

subgroups, based upon sample biochemistry. The data analysis techniques used are described in section 2.4 below.

Potential applications include use as an ‘optical biopsy’ technique for *in vivo* tissue diagnosis or to guide therapy, as a ‘digital staining’ method to assist a histopathologist in analysing a sample, or as an entirely automated process for histopathology classification.

Developments in vibrational spectroscopic applications in the oesophagus are reviewed below, but it should be noted that these techniques are also being investigated to accurately identify a range of pathologies in a number of other tissue sites, including stomach (Bergholt et al. 2013; Bergholt, Zheng, Lin, Ho, Teh, Yeoh, Yan So, et al. 2011; Ito et al. 2013; Li et al. 2013), colon (Lloyd et al., 2012), breast (Kast et al., 2007; Kong et al., 2014), lymphoma (Isabelle et al. 2008), lung (Gao et al. 2012; Huang et al. 2003; Krafft and Sergo 2006; Pijanka et al. 2013; Short et al. 2008), cervix (Bazant-Hegemark et al. 2008; Lyng et al. 2007; Martinho et al. 2008.; Robichaux-Viehoever et al. 2007; Wood et al. 2004), brain (Gajjar et al. 2012; Ji et al. 2013; Toms et al. 2006), head and neck (Das et al. 2006; Harris et al. 2010; Menzies et al. 2013; Singh et al. 2012; Stone et al. 2000), skin (Hammody et al. 2008; Nijssen et al. 2007; Tfaili et al. 2012; Yu et al. 2013), and urological applications (Baker et al. 2009; Barman et al. 2012; Crow, Barrass, et al. 2005; Crow, Molckovsky, et al. 2005; Derenne, Gasper, and Goormaghtigh 2011; Draga et al. 2010; Grimbergen et al. 2010; Stone et al. 2007).

2.2.2. Clinical applications of Raman in the oesophagus

2.2.2.1. Raman tissue mapping: towards automated histology

Raman spectroscopy has shown promise in differentiating pathological subtypes in the oesophagus, with much of the initial work performed in vitro. In an analysis of snap-frozen oesophageal biopsy samples, Kendall et al. showed the ability to discriminate between 8 pathological groups: normal squamous, 3 subtypes of Barrett's oesophagus (intestinal metaplasia, cardiac and fundic type mucosa), high-grade dysplasia, adenocarcinoma, squamous dysplasia and squamous cell carcinoma (Kendall et al. 2003). When analysed as a 3-subset model (normal vs. Barett's vs. dysplasia/adenocarcinoma), they achieved a high level of agreement with consensus pathological opinion (k=0.89) (Kendall et al. 2003). This level of agreement compared favourably with that of an independent pathologist analysing the same samples against consensus pathology (k=0.76), demonstrating the importance of consensus opinion, and underlining its value in training spectral models.

Raman spectral mapping of thin oesophageal tissue sections has been used to characterise the biochemical changes underlying the spectral differences. Shetty et al. used pseudo-colour principal component score maps to identify altered cellular constituents associated with malignancy (Shetty et al. 2006). Dysplastic glandular tissue showed higher levels of DNA, actin and oleic acid, whereas normal squamous tissue had a higher glycogen content.

Rapid mapping of tissue sections using Raman spectroscopy has the potential to be used as an automated histopathology tool. Hutchings et al. showed that mapping could be performed in clinically applicable timescales e.g. 2mm

diameter section over 30-90 minutes, and that this was sufficient to discriminate pathology (Hutchings et al. 2009).

2.2.2.2. Raman probe: towards an 'optical biopsy'

As a hollow organ which can be visualised readily with endoscopy, much work in the oesophagus, and elsewhere in the gastro-intestinal tract, has been aimed at the development of a fibre-optic Raman probe for *in vivo* use, with several probes having been designed to be compatible with medical endoscopes. The potential advantages of Raman as an 'optical biopsy' tool are outlined in Table 2-1 below.

Table 2-1 Putative advantages of an 'optical biopsy'

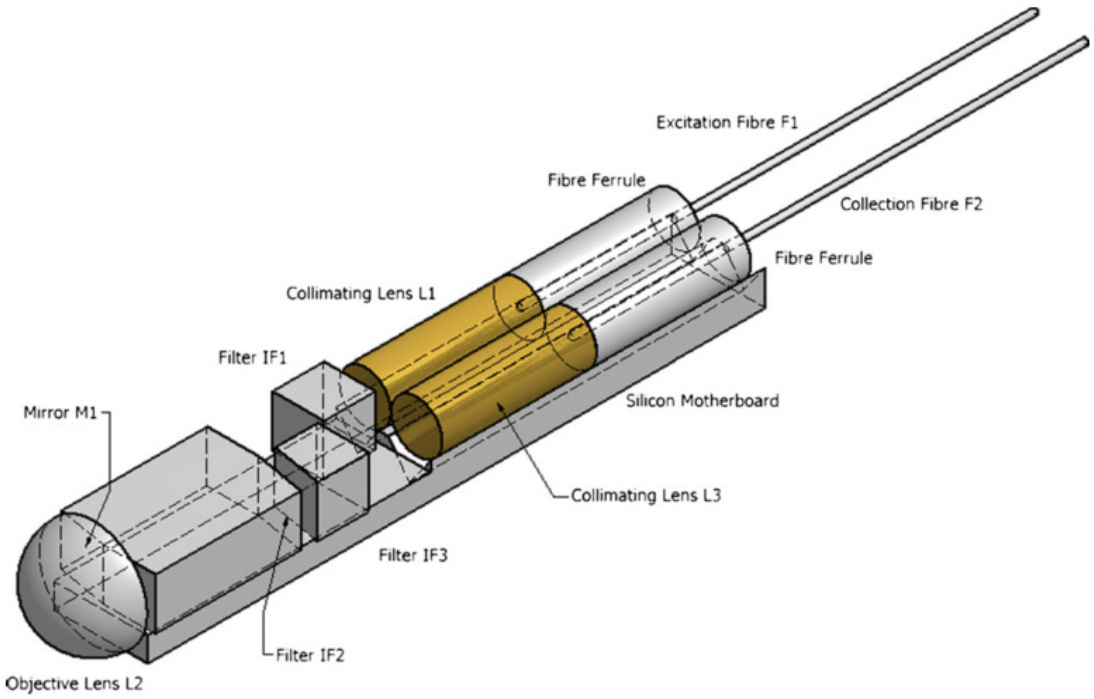
Advantages of optical biopsy	**Clinical benefits**
Objective evidence for diagnosis	Avoids subjectivity of histopathological assessment
Biochemical changes may precede morphology	Earlier detection of pre-malignant changes
Reduce need for tissue biopsy	Avoids risk of bleeding, risk of perforation of hollow viscus, cost of tissue processing, cost of analysis, delay to diagnosis.
Enable targeted cancer resection	Ensure resection margins are clear intra-operatively, avoiding need for re-excision or uncertainty over completeness of excision.

Reproduced from Old et al. 2014.

There are a number of design and manufacturing challenges which must be overcome to achieve this in the oesophagus. The probe must fit the instrument channel of a gastroscope, the internal diameter of which measures 2.8mm for a standard endoscope (though this can vary from 2.0-3.8mm) (Olympus medical 2013; Varadarajulu et al. 2011) and a scope length of 92.5-110cm (Varadarajulu et al. 2011). The probe must be sufficiently durable to withstand manipulation through articulation of the endoscope, and resistant to degradation by both biological secretions and processes of disinfection. Background scattered light must be filtered prior to spectrometer detection, requiring multiple filters. The acquisition time must be short to ensure a stable reading, and to allow the examination to proceed in a clinically appropriate timescale.

Using a custom-built Raman probe, designed to be compatible with standard medical endoscopes, Almond et al. were able to demonstrate accurate recognition of pathology subtypes (Almond et al. 2012, 2013). A schematic diagram of the probe, along with a picture of the packaged probe, is shown in Figure 2-5. Using spectra acquired in 1 second, with a 3 group analysis (normal squamous cells vs. Barrett's oesophagus/LGD vs. high grade dysplasia/adenocarcinoma) gave sensitivity and specificities for normal squamous cells of 87% and 96% respectively, Barrett's/LGD 72% and 91%, HGD/adenocarcinoma 86% and 88%. Using a 2 group model with spectra acquired over 5 seconds the detection rate of HGD/adenocarcinoma remained 86%, but the specificity was greatly improved at 98%.

A



B

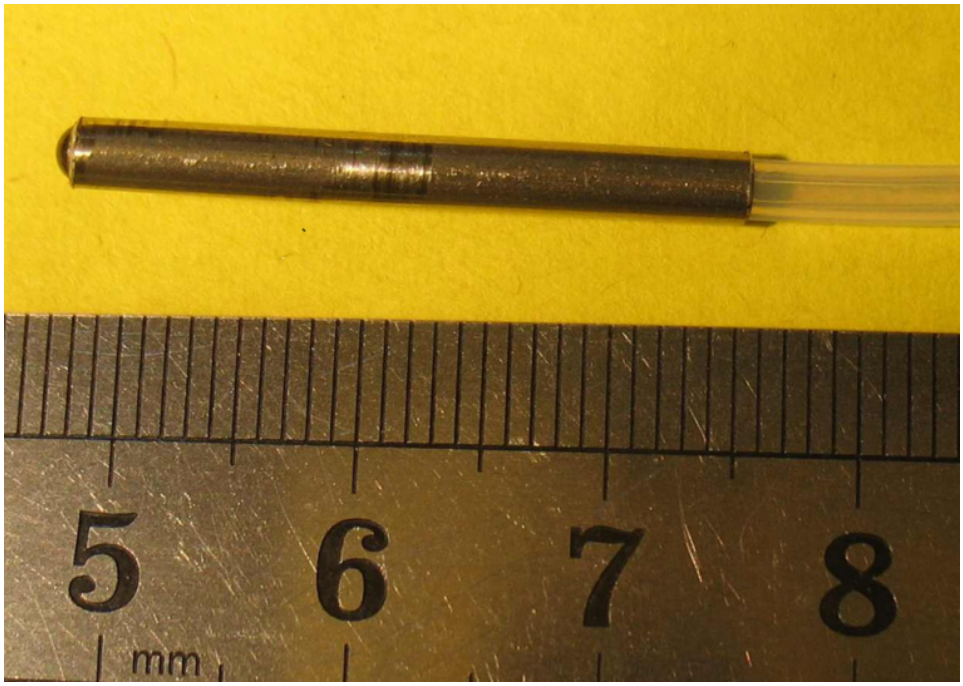


Figure 2-5 A) Schematic diagram of Raman probe. B) Packaged probe. Reproduced from Day et al. 2009.

A number of groups have trialled Raman probes *in vivo* in the oesophagus. A preliminary study in 2000 demonstrated the safety and feasibility of using such probes in the oesophagus, using an Enviva Visionex probe (Shim et al. 2000). This group measured 400 spectra in 20 patients using 5 second acquisition times, although discrimination of pathology was poor (Shim et al. 2000).

In 2005 the same group used a custom-built probe to measure 192 spectra in 65 patients with Barrett's oesophagus, using 5 second acquisition times. They were able to discriminate Barrett's, LGD and HGD/adenocarcinoma with accuracy 88%, 81% and 92% respectively. When analysed using a 2 subset classification, they reported sensitivity of detecting HGD/adenocarcinoma as 88% with a specificity 89% (Wong Kee Song and Wilson 2005; Wong Kee Song et al. 2005).

More recently, Bergholt et al. have used a Raman probe to identify cancer in the oesophagus (Bergholt, Zheng, Lin, Ho, Teh, Yeoh, So, et al. 2011). Whilst most of the early work by this group has been in the stomach, they have recently begun using their Raman probe in the oesophagus. Initially they measured 263 oesophageal spectra in 80 patients (33 with cancer, of whom 26 were adenocarcinoma and 6 were squamous cell carcinoma) with acquisition times of 0.4-0.5 seconds, reporting identification of cancer in the oesophagus with sensitivity 91% and specificity 94% (Bergholt, Zheng, Lin, Ho, Teh, Yeoh, So, et al. 2011). Prior to commencement of our project this probe had not been used to identify Barrett's oesophagus.[1]

As a 'point-imaging' technique which provides biochemical information about tissues, there is potential for Raman spectroscopy to be used in combination

[1] This group has recently presented data from testing the probe in patients with Barrett's oesophagus – this is discussed further in the Summative Discussion.

with other imaging modalities, either using initial ‘wide-field’ imaging to identify areas of interest to be probed by Raman, or with modalities providing different types of information about the tissue. Raman probes have been used in conjunction with a trimodal imaging system which combines high-resolution white light endoscopy, autofluorescence imaging and narrow-band imaging, in a system known as trimodal imaging (Bergholt, Zheng, Lin, Ho, Teh, Yeoh, So, et al. 2011). This multimodal system aims to increase detection of suspicious lesions using autofluorescence imaging which has a high detection rate but is associated with a high false positive rate (Boerwinkel et al. 2013). Narrow band imaging, which is more specific for dysplasia than AFI (Urquhart, DaCosta, and Marcon 2013), can then identify areas for biopsy, or in combination with Raman, ‘optical biopsy’.

Some groups have used optical coherence tomography (OCT) in conjunction with Raman for certain cancer types. Optical coherence tomography provides information about tissue architecture, and where Raman identifies malignant changes through biochemistry, OCT could be used to measure depth of local invasion, thus informing decisions about treatment e.g. whether endoscopic resection would result in a compete excision or more extensive surgery is required. Although not yet performed in the oesophagus, *in vivo* feasibility work has been done on human skin (Patil et al. 2011), and on *ex vivo* tissue samples in the colon (Ashok et al. 2013).

2.2.3. Advanced Raman techniques

2.2.3.1. Coherent Anti-Stokes Raman Spectroscopy (CARS)

Coherent Anti-Stokes Raman spectroscopy (CARS) is a non-linear variant on Raman microspectroscopy, which provides a signal that is orders of magnitude

stronger than the spontaneous Raman signal from biological tissue, and consequently can be acquired far more rapidly (Tu and Boppart 2014).

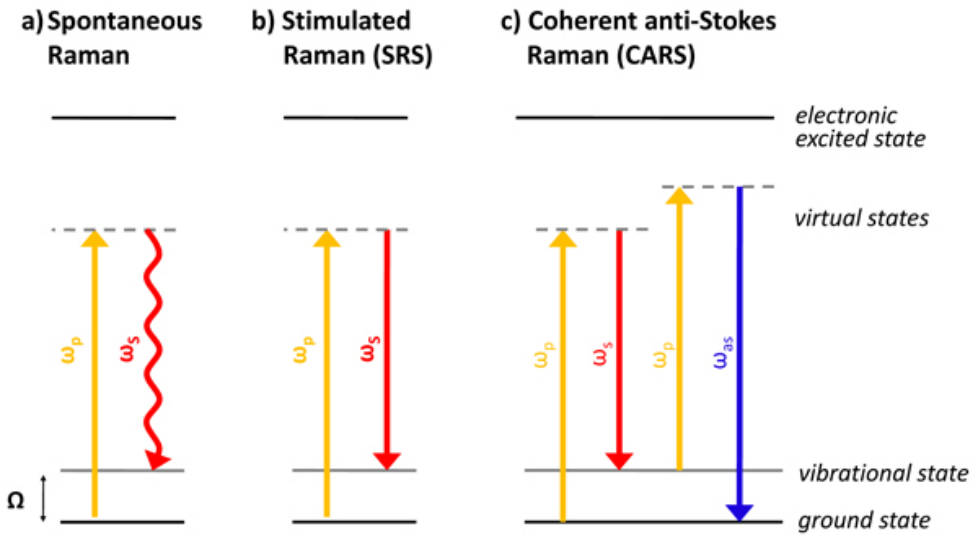


Figure 2-6 Comparison of energy changes in a) spontaneous Raman, b) Stimulated Raman and c) CARS. Reproduced from Harvard University Department of Chemistry and Chemical Biology 2015.

Spontaneous Raman involves excitation from incident pump photons resulting in a scattered photon: if initially at ground energy state the scattered photon is at Stokes frequency (ω_s). If both a pump photon at frequency ω_p, and a further photon at Stokes frequency (ω_s) are incident on the sample, and the frequency difference matches the vibrational frequency (Ω) of a Raman-active molecule in the sample (i.e. $\Omega = \omega_p - \omega_s$), stimulated excitation occurs (known as stimulated Raman scattering, or SRS). If a further pump (or 'probe') photon (ω_{pr}) is then incident on the sample, a photon at Anti-Stokes frequency (ω_{as}) is then emitted. This four-wave mixing process forms the basis of CARS (see Figure 2-6) (Harvard University Department of Chemistry and Chemical Biology 2015.; Tu

and Boppart 2014). A schematic diagram of the instrumentation is shown in Figure 2-7.

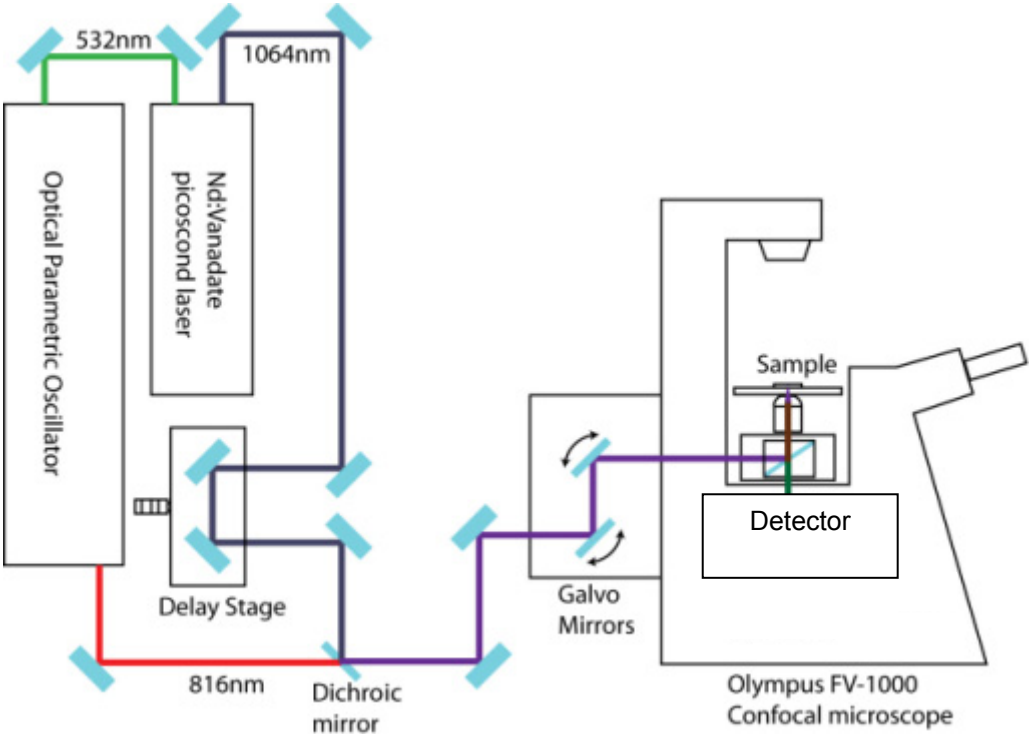


Figure 2-7 Schematic diagram of CARS imaging. A picosecond laser emits beams of different wavelengths: one passes through an optical parametric oscillator, the other through a delay stage before their paths rejoin to be focussed onto the sample using an inverted microscope. Adapted from Lei et al. 2011.

The resonant vibrations produced by CARS lead to a signal greater than 5 orders of magnitude stronger than spontaneous Raman (Begley, Harvey, and Byer 1974). This feature of CARS is its chief advantage, with the result that acquisition times can be drastically reduced. This provides the possibility of real-time, label-free imaging, capable of video-speed analysis. Applied to histological sections, mapping can be achieved much faster than with

spontaneous Raman, where a map that may take many hours can be acquired in just a few minutes. The major advantages of CARS over spontaneous Raman are outlined below (Evans and Xie 2008):

1. The signal is orders of magnitude stronger than spontaneous Raman.
2. Speed – the stronger signal allows much faster acquisition times.
3. It measures intrinsic molecular vibrations so no extrinsic label (e.g. antibody biomarker, nanoprobe etc.) is needed.
4. There is no interference from autofluorescence with the Anti-Stokes signal from CARS.
5. The nonlinear photon beams generate CARS signal at a focal spot, allowing 3D tissue interrogation at varying depths; at near-infrared wavelength depths of up to 0.4mm may be achieved.
6. CARS causes minimal sample photodamage (Evans et al. 2005).

However there are a number of limitations with this technique. CARS measures predominantly in the high wavenumber range above 1800cm^{-1}, and does not include most of the 'Fingerprint region' which is most useful in discriminating complex biological substrates. For this reason, much work to date has been on imaging lipids, which give a signal in this high wavenumber region. With such strong signal, video-rate imaging is possible for lipids (Evans et al. 2005).

A further limitation of CARS is the method of measuring one frequency at a time. Thus for mapping applications, maps are built up from multiple sequential measurements, each at a different frequency, but for rapid imaging modalities, only a few discrete frequencies can be studied.

Other major challenges with CARS include the non-resonant background, which makes interpretation of spectra very difficult (Krafft, Dietzek, and Popp 2009), and the limitation that CARS is unable to provide a quantitative measure of a sample, since the resonant signal is not proportional to sample content (Tu and Boppart 2014). Finally, an obvious barrier to clinical translation is the cost and complexity of the instrument set up (Tu and Boppart 2014).

2.2.3.2. Clinical applications of CARS

A number of groups have investigated the use of CARS in mapping histology samples as a diagnostic tool for pathology. To date these have largely been proof of concept studies that demonstrate the ability to distinguish salient features of disease that may aid diagnosis.

CARS has been used in the gastrointestinal tract to discriminate between cell types using samples from human colon (Krafft, Ramoji, et al. 2009), by identifying the mucin content of goblet cells in colonic crypts.

Discrimination between cancer subtypes has been shown in a small, exploratory study by Gao et al. using a lung cancer model (Gao et al. 2012). They reported excellent results for differentiation between small cell and non-small cell tumours, and for discrimination between adenocarcinoma and squamous cell carcinoma.

Accurate identification of brain tumours has been demonstrated in an orthotopic mouse model, and confirmed in human *ex vivo* tissue (Uckermann et al. 2014).

Table 2-2 Comparison of the characteristics of CARS required for clinical spectroscopy and clinical imaging (adapted from Tu and Boppart 2014).

	Clinical spectroscopy	**Clinical imaging**
Raman shift of interest	Broadband continuous spectrum	Single/few discrete frequencies
Tissue samples	*Ex vivo*, thin (<100μm) section	*In vivo* tissue, thick
Instrumentation	Microscopy/micro-spectroscopy based on free-space optics	Fibre-optic based devices
Signal detection	Array detector	Single detector
Operator	Laboratory technician/Researcher	Clinical staff
Environment	Clinical laboratory	Point-of-care
Representative applications	Research, molecular histopathology, Raman biomarker identification	Optical biopsy, intra-operative assessment, disease/treatment monitoring

CARS can also be used to highlight subcellular structures. El-Mashtoly et al. (El-Mashtoly et al. 2014) used a cultured pancreatic cancer cell line to delineate several cell organelles including the nucleus and nucleolus, mitochondria, Golgi apparatus, endoplasmic reticulum and intracellular lipid droplets.

The potential for video-rate imaging has prompted development of a fibre-optic instrument to allow *in vivo* endoscopic CARS. This technique has potential for clear advantages over spontaneous Raman in that it could be used as a label-free wide-field imaging technique. One of the most exciting prospects recently investigated is the coupling of CARS technology with scanning fibre

endoscopes: a single fibre is driven at its mechanical resonance frequency by a piezoelectric actuator to generate a 2D image (Balu et al. 2010; Gao et al. 2012). Much work remains to refine this technique but future developments in fibre-optic CARS are eagerly awaited. A comparison of the characteristics required as a clinical imaging tool versus a spectroscopic (i.e. mapping) tool is shown in Table 2-2.

2.2.3.3. Other advanced Raman techniques

Standard Raman spectroscopy only has a penetration depth of a few hundred microns, thus restricting laser probing to the surface or near-surface area of samples (Matousek and Stone 2009). The development of spatially offset Raman spectroscopy (SORS) enables spectral measurements from areas deep into the sample (Matousek 2009). The technique is based on the principle that photons migrate. Spectra are measured through the collection of scattered photons from surface areas on a sample that are laterally offset at a distance from the excitation laser point of incidence. A range of spectra are generated with varying intensities relating to the surface and subsurface layers. As collection becomes further away from the point of laser excitation, the spectral contributions from the top layer of the sample decrease faster than those from deeper layers. This is due to the likelihood of greater lateral migration of photons generated deeper in the sample before they are emitted from the sample surface. Through processing of the dataset using multivariate analysis, pure spectra can be produced of the individual layers of the sample (Matousek et al. 2005). Clinical uses of deep Raman techniques may include applications

in urology, bone and breast disease (Baker et al. 2007; Matousek and Stone 2013; Matousek 2009; Matousek et al. 2005).

There are a number of other advanced Raman techniques including hyper-Raman, stimulated Raman spectroscopy (SRS), surface enhanced Raman spectroscopy (SERS), and tip-enhanced Raman spectroscopy (TERS), that are beyond the scope of this review.

2.3. Infrared Spectroscopy

2.3.1. Infrared spectroscopy of biological samples

Infrared spectroscopy depends upon the absorption of incident radiation by the sample tissue, with measurement of the resultant energy loss (Isabelle, Stone, and Barr 2008). For a photon of light to be absorbed by a molecule, the photon's energy must correspond to the energy gap between the molecule's current state and a higher vibrational energy state, with a consequent loss of energy from the incident light. The vibrational frequency of a molecule and thus the probability of absorption are influenced by adjacent molecular bonds, and consequently the wavelength (or wavenumber, 1/wavelength) of absorption bands provides information about the chemical composition of the tissue (Barth 2007). By measuring the attenuation (reduction in intensity) of the transmitted light after passing through a sample, an absorption spectrum can be produced.

Chemical bonds between atoms undergo various forms of vibration; most molecular vibration occurs at energies corresponding to the infrared region of the electromagnetic spectrum, and consequently infrared radiation can produce excitation of molecular bonds (Haris and Chapman 1992). Vibrations that cause

changes in the dipole moment of the molecule can be measured using infrared spectroscopy (Haris and Chapman 1992), with the infrared spectrum measuring how much energy has been absorbed at a particular wavelength. Thus for a large molecule such as a protein, there are many possible vibrations and a complex spectrum will be produced.

The potential of infrared (IR) spectroscopy to provide information on the secondary structure of proteins was first described by Elliott and Ambrose in 1950 (Elliott and Ambrose 1950) but it was not until the advent of Fourier Transform spectrometers in the 1980s that IR spectroscopy became widespread (Van de Weert, Hering, and Haris 2005). Fourier transform spectrometers enable rapid data collection and high light intensity at the detector to achieve a high signal to noise ratio (Barth 2007); virtually all modern IR spectroscopy is performed using Fourier Transform Infrared (FTIR) (Kendall et al. 2009).

In addition to protein conformation, FTIR spectroscopy has the ability to characterise the structure of lipids, nucleic acids and carbohydrates (Kendall et al. 2009; Quaroni and Zlateva 2011). In the context of clinical diagnostics, the technique has also been applied to more complex biological substrates such as tissue samples, body fluids and cell cultures. In this setting, since many bond vibrations are common to many biomolecules, FTIR does not measure specific molecules, but rather is used to quantify classes of molecules e.g. glycogen, protein, nucleic acid present in a sample (Mackanos and Contag 2010). Thus for a given tissue sample, FTIR generates a characteristic spectrum of infrared absorption peaks and consequently a unique ‘biochemical fingerprint’ based on

the overall biochemical composition of the tissue, with varying contributions from each type of bond vibration (see Figure 2-8).

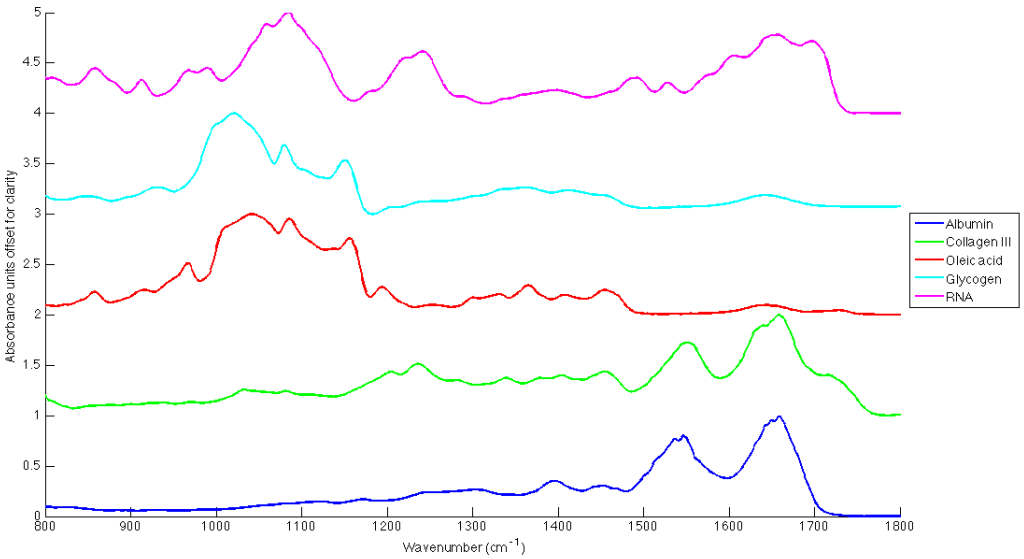


Figure 2-8 Infrared spectra of selected key biomolecules – infrared absorption provides characteristic spectral peaks based on whole sample biochemistry, allowing classification of complex molecular structures or biological substrates.

This ability to probe the biochemical characteristics of tissues enables the use of FTIR to discriminate between cells in different stages of their life cycle and proliferation, identify metaplastic change in cell type, and recognise cancerous changes (Boydston-White et al. 2006; Kendall et al. 2009; Romeo et al. 2006).

Advantages of infrared spectroscopy include a large range of applications from small proteins to tissue samples, a short measuring time, a small amount of sample required and relatively low costs (Barth 2007).

2.3.2. Instrumentation

Historically, the use of FTIR *in vivo* has been limited because of the challenges posed by the high water content of most biological tissues, as water is highly absorptive in the mid-infrared range in which FTIR operates (Kendall et al. 2009; Mackanos and Contag 2010). Consequently, much of the work has been performed on excised tissue specimens, which must be sectioned and dried before sampling. However, advances in instrumentation and sampling techniques are reducing the need for tissue preparation, and addressing the challenges of *in vivo* application.

FTIR spectrometers use a radiation source with the optical path passing through an interferometer to provide a calibrated interferogram (see Figure 2-9) (Bhargava 2012); this is the Fourier transform of the spectrum. In transmission mode spectroscopy, after passing through the sample the signal is measured using a single detector, and a computer then performs a second Fourier transform to convert the measured data back into an infrared transmittance spectrum over the measured wavelengths (Barth 2007).

Reflectance techniques can be used instead of transmission, in which the infrared radiation reflected from the surface of the sample is measured, or in transmission-reflection (or 'transflectance'), where radiation passes through a sample then hits a reflective surface, passes back through the sample and to a detector. This requires more complex post-processing to correct for misalignment of spectral peaks due to refraction within the sample.

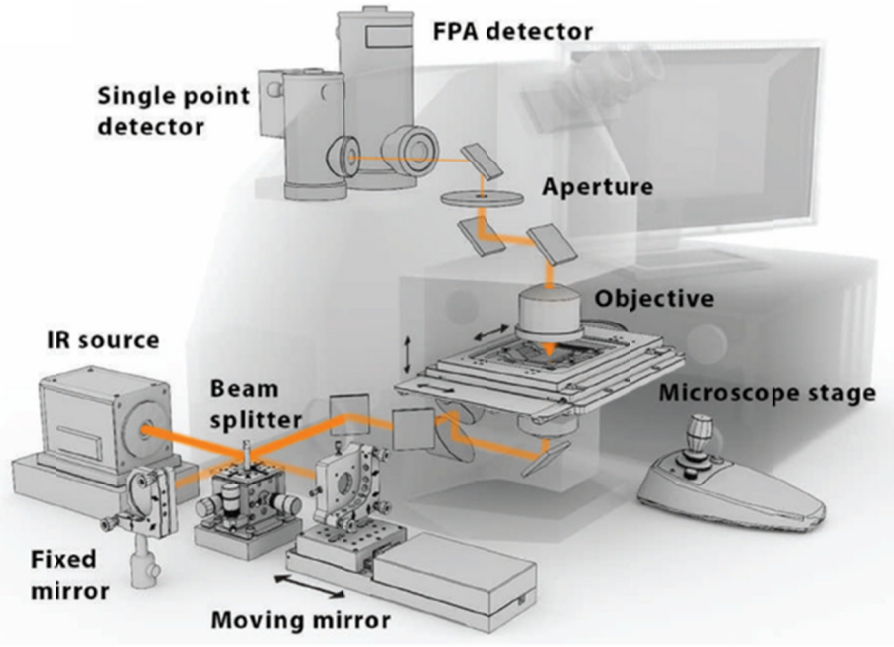


Figure 2-9 Typical layout of a Fourier-transform spectrometer (reproduced from Bhargava 2012).

Attenuated Total Reflection (ATR) spectroscopy is a technique which overcomes some of the limitations of transmission or reflectance mode FTIR spectroscopy. In ATR, the IR beam is guided through an IR transparent crystal placed in contact with the sample. The crystal has a high refractive index, and light is directed towards the sample interface at incident angle such that total internal reflection occurs at the crystal surface, and an evanescent standing wave penetrates the sample (Barth 2007; Kendall et al. 2009). This evanescent wave has the same frequency as the incoming light but exhibits exponential decay with distance from the interface. Thus the evanescent wave may be differentially absorbed by the tissue sample as determined by its biochemical composition, but sampling only over a very superficial distance. The sample penetration depth can be precisely controlled by varying the angle of incidence, and the ratio of refractive indices of the crystal and sample (Barth 2007; Kendall et al. 2009).

This technique has the major advantage of requiring little or no tissue preparation, as the difficulties of sample thickness and water absorption are overcome. ATR also enables higher spatial resolution than is possible with transmission IR, with micro-ATR capable of spatial resolution of around 4-5μm (Kazarian and Chan 2013). Disadvantages include the need for contact with the sample and the potential for cross-contamination of the instrument as a result.

The development of endoscope-compatible fibre-optic probes has opened the door to the use of FTIR in numerous hollow organs *in vivo*. Flexible infrared light guides have been constructed from a number of materials, with silver halide preferred for *in vivo* use since it is transparent in the mid-IR, non-toxic, flexible, non-hydroscopic and stable (Mackanos and Contag 2010). The ability of such probes to discriminate pathology from normal tissue has been demonstrated in *ex vivo* samples from multiple sites, including oral mucosa, colon and bladder (Mackanos and Contag 2010).

The ability to correct for the absorption of paraffin has allowed FTIR to be used on formalin-fixed tissue samples, and this provides the potential to validate sample spectra using extensive tissue banks (Nallala et al. 2012).

A comparison of some of the key features of Raman and FTIR spectroscopy is included in Table 2-3.

Table 2-3 Comparison of Raman and FTIR spectroscopy (adapted from Kendall et al., 2009).

	Raman	**FTIR**
Wavenumber range	400-4000cm^{-1}	800-4000cm^{-1}
Light source	Monochromatic laser	Polychromatic light source
Detection method	Inelastic light scattering	Mid-infrared absorption
Molecular conditions for activation	Changes in polarizability	Changes in dipole moment
Molecular bond sensitivities	Non-polar bonds: C-C bonds including double/triple/aromatic rings	Strong polar bonds: hydroxyl (OH), carbonyl (CO) and amide bonds
Signal-to-noise ratio	Lower	Higher over same timescales
Spatial resolution	Higher (1μm)	Lower (5μm). Synchroton sources 2-5μm
Imaging/mapping modes	Point raster mapping, line and rapid readout mapping. Fast Raman imaging, ultrafast confocal Raman imaging, wide-field imaging, LCTF Raman imaging.	Rapid scan imaging using focal plane/linear array detectors.
Enhanced techniques	SORS, CARS, SERS, TERS	ATR
Sampling challenges:		
Contact with sample	Required for *in vivo* probes	Not for transmission; necessary for ATR
Sample thickness	Point measurements from thick sections, 5-20μm for mapping	Thick samples (>15 μm) can cause spectral saturation
Water	Weakly scattered	Strong absorption
Paraffin	Strong peaks in fingerprint region	Visible in fingerprint region, can be overcome by spectral subtraction.

LCTF Liquid Crystal Tunable Filter; SORS Spatially Offset Raman Spectroscopy; CARS Coherent Anti-Stokes Raman Spectroscopy; SERS Surface Enhanced Raman Spectroscopy; TERS Tip-Enhanced Raman Spectroscopy; ATR – Attenuated Total Reflectance.

2.3.3. FTIR in the oesophagus

2.3.3.1. FTIR for oesophageal histology

The potential of FTIR to provide a rapid, objective and non-destructive histopathological analysis has been recognised for several decades (Quaroni, Zhao, and Casson 2009). In part because of the challenges posed by the strong absorption by water in the mid-infrared region, FTIR is less well-suited to use *in vivo*, and in the oesophagus work has focused on developing a tool as an adjunct to histopathology.

In 2007, Wang et al. used ATR to study 98 fresh oesophageal specimens from 32 patients (Wang et al. 2007). This group found that DNA, protein, glycoprotein and glycogen account for most absorption in the 950 to 1800cm^{-1} region. Using a system of partial least squares to classify disease states, they could identify dysplasia with sensitivity 92% and specificity 80% (Wang et al. 2007).

The feasibility of FTIR in formalin-fixed, paraffin-embedded oesophageal tissue sections has also been demonstrated (Quaroni and Casson 2009). Spectral differences based on accumulation of glycoproteins were used to identify Barrett's oesophagus. Synchrotron data showed that the source of these glycoproteins was the goblet cells which are characteristic of intestinal metaplasia in Barrett's oesophagus (Quaroni and Casson 2009).

Amrania et al. have attempted to develop an adjunct for histopathologists that can rapidly scan tissue sections, and then produce a composite, 'digitally stained' image which could make interpretation faster and more accurate (Amrania et al. 2012). This utilises FTIR to provide a measure of DNA to

cytoplasmic ratio across the slide, which is then used to create a false-colour image which can be viewed by the histopathologist.

2.3.3.2. Infrared spectral cytopathology

Applications of spectroscopy to pathology have traditionally been based on tissue diagnosis, and compared against histology classification: the use of FTIR to analyse cells has been termed Infrared Spectral Cytopathology (SCP). Initial work on cells using FTIR measured whole cell pellets (Wong et al. 1991), but advances in micro-spectrometers have made it possible to use FTIR to measure spectra from individual cells (Diem et al. 2002, 2015). The biochemical composition of the cell measured by FTIR can reflect subtle changes in the cell, such as stage of the cell cycle or disease. As with spectroscopic tissue analysis described above, chemometric methods can be used to classify cells based on their spectra: this can be used to characterise progression through the cell cycle, cell type or pathological sub-group (Bird et al. 2008; Diem et al. 2015; Jimenez-Hernandez et al. 2013; Papamarkakis et al. 2010; Schubert et al. 2010) .

Nonetheless, there are a number of challenges posed by spectral interrogation of individual cells, not least the fact that signal from such a small sample is inevitably weak, and signal to noise ratios are consequently higher than for tissue measurement.

Another major challenge is the Mie-type scattering effects which occur due to the cells or cell components (particularly the nucleus) being of similar size to the incident wavelength. This causes an oscillation in the spectral baseline and can

affect the position and intensity of spectral peaks (Bassan et al. 2009). This effect is a potential confounder in any subsequent interpretation of the affected peaks. A correction algorithm (extended multiplicative scatter correction, EMSC) has been developed to account for these spectral distortions (Kohler et al. 2008).

A further scattering effect is the 'dispersion artefact', which can alter the intensity and position of the Amide I peak at around 1655cm^{-1}, which is frequently an important diagnostic marker in cell spectroscopy, and also causes pronounced effects on peaks to the high wavenumber side of the Amide I peak (Bassan et al. 2009, 2010; Miljković, Bird, and Diem 2012). Bassan et al. demonstrated that the dispersion artefact was caused by resonant Mie scattering – i.e. scattering when there is simultaneous absorption, since Mie scattering depends on the refractive index which changes on passing through an absorption resonance. Algorithms have been developed to correct for this artefact (Bassan et al. 2009; Miljković et al. 2012): computational data pre-processing steps will be discussed further in the appropriate methodology section.

The development of such computational tools has minimised the effects of spectral distortion and enabled progress to be made in investigating clinical applications of SCP.

2.3.3.3. Clinical applications of SCP

If SCP could be harnessed in clinical practice, it could have a role in cytological diagnosis for a range of disease processes. Potential advantages over

conventional cytology include automation and reduction in clinical staff resources, the potential for objective measurement based on cellular biochemistry, every cell is measured so no cell can be missed through human error, and the ability to detect biochemical change that precedes morphology. It also has the advantage over histology of offering a relatively less invasive method of diagnosis, as tissue biopsy is not required and cells can be acquired by a variety of means e.g. in bodily excretions such as urine, or via superficial scrapings from an accessible hollow viscus, plus the ability to sample large regions simultaneously without taking multiple biopsies.

In countries such as the UK that have introduced cervical screening programmes, cervical smear cytology frequently forms the largest portion of a cytology departmental workload. There has been much work previously to introduce automation for cervical screening based on image software detection of DNA content and enlarged nuclei (Biscotti et al. 2005), which led to a large randomised controlled trial in the UK comparing manual reading against automated or paired (automated then manual) reading (Kitchener et al. 2011). Unfortunately, this showed a lower sensitivity in the automated group and an inconsequential rise in specificity, so its use has not been recommended.

Consequently cervical cytology is one of the first areas that has been explored by SCP, and a number of groups have published work in this field. The first such study by Wong et al. measured signal from large numbers of cells in a cell pellet rather than individual cells, but were able to distinguish differences between normal, dysplastic and cancerous samples, using contemporaneous stained slides as a reference standard (Wong et al. 1991). Further work on cell pellets by the group led by Max Diem, demonstrated the ability to differentiate

stages of maturation of cervical epithelial cells (Chiriboga et al. 1998). Later work by the same group showed differences could be identified between normal and cancerous cultured cells at a single cell level (Diem et al. 2002). Analysis of individual cells has continued to be the focus of more recent research. Schubert et al. demonstrated the use of SCP to discriminate normal squamous cells from low grade intraepithelial lesions in a group of 12 patients, and went on to show the ability of SCP to correctly identify HPV infection in samples from 2 patients with HPV and 3 without (compared to HPV DNA testing) (Schubert et al. 2010).

Gajjar et al. recognized that inaccuracies in manual cytology reporting could introduce a teaching error into the training model, and consequently attempted to improve the 'gold standard' against which to construct their dataset (Gajjar et al. 2014). They took contemporaneous biopsy results and used histology, rather than cytology, as the gold standard. They demonstrated that SCP was superior to manual cytology in identifying changes when compared to histology.

A further study applied SCP to cells lining the oral mucosa (Papamarkakis et al. 2010). This demonstrated that spectral differences between cells from different areas of the oral cavity could be identified in 6 healthy volunteers, and that these samples could be separated from a further small group of 3 samples showing cancerous changes.

SCP has also been shown to accurately discriminate between different cell types from the urinary tract in urine samples (Bird et al. 2008). This is a first step towards a possible role for SCP in bladder cancer detection from urine samples.

2.4. Interpreting spectra: chemometrics and data analysis techniques

2.4.1. Data pre-processing

Prior to spectral analysis, a number of techniques may be applied as pre-processing steps to correct for measurement artefacts or to allow comparison of spectra, whilst minimising systematic or random errors in the measured data. Normalisation of spectra reduces the effect of differences in intensity of the measured spectra (which may occur with e.g. different tissue thickness, or varying incident power). In order to highlight differences between spectra, mean centring may be used, whereby the mean spectrum of the entire dataset is subtracted from each spectrum.

Correction algorithms for scattering effects (EMSC and resonant Mie scattering) are discussed above in section 2.3.3.2. EMSC is also used more broadly to correct for any physical effect that causes differences in the measured spectra, using fitting techniques such as least squares may be used to remove differences in the baseline spectrum (Lieber and Mahadevan-Jansen 2003), correct for contributions from the objective lens, or correct for the presence of contaminants in the tissue e.g. paraffin or OCT used in tissue preparation (Nallala, Lloyd, and Stone 2015).

2.4.2. Principles of chemometrics

Biological spectra are composed of the spectral contributions of many constituent molecules: as a result of this they contain a wealth of information about the composition of the sample, but this makes visual interpretation of such spectra challenging. Furthermore, the differences between spectra from

similar biological samples (e.g. different pathology subtypes from the same tissue type) may be subtle, and not readily apparent simply by looking at the spectra themselves. Thus further analysis is required to characterise the measured spectra, the contributions from certain biomolecules with known spectral peaks, and any consistent differences between sample groups that represent a true biochemical difference. These differences between samples can then be used to develop classification models which predict some property (e.g. pathology group or cell type) of new test samples.

A number of statistical approaches may be used to identify the spectral features that represent the biochemical differences between samples. Large databases of spectra may be used to form a training dataset, with chemometric pattern recognition techniques then applied, such as prinicipal component analysis (PCA), linear discriminant analysis (LDA), partial least squares (Lloyd et al. 2013) and support vector machines (Sattlecker et al. 2010). If such techniques are applied they must then be validated using an independent test set or another appropriate method (e.g. leave one sample out cross validation, LOSOCV) to give an accurate measure of diagnostic performance. The techniques applied in this thesis (PCA, LDA and LOSOCV) are discussed further below.

Another method of analysis is to use clustering techniques, such as hierarchical cluster analysis (HCA), self-organising maps (SOM) and k-means clustering (Lasch et al. 2004; Lloyd et al. 2012). These provide a visual representation of the grouping of data.

2.4.3. PCA, LDA and LOSOCV

Multivariate analysis of large datasets is challenging due to the huge volume of data. PCA is a method of describing the variance within a dataset that uses correlation to achieve data reduction (Miller 2010). In the situation with only 2 sources of variation, there will be 2 principal components, as shown in Figure 2-10. It can be considered a rotation of the axes, where principal component 1 (PC1) represents the maximum source of variation, and PC2 then represents the next greatest source of variation, with PC2 being at right angles to (i.e. orthogonal to) PC1. Applied to spectra with many variables, it identifies the maximal source of variance between the spectra with each sequential component: thus PC1 accounts for the largest difference between the spectra, followed by PC2, then PC3 etc.

Analysis of variance (ANOVA) can be applied to the principal components to identify those PCs which describe the greatest sources of variance. These PCs may carry the most useful information for subsequently building classification models.

PCA assumes no prior knowledge about the dataset or any groupings within it, and is thus considered an ‘unsupervised’ technique. Providing the data analysis model with ‘group membership’ information for each spectrum can improve classification performance, and is known as a ‘supervised’ technique.

Linear discriminant analysis (LDA) is one such supervised technique that incorporates information regarding the group each datapoint comes from. It uses this information to identify differences between the groups to enable classification models to be developed.

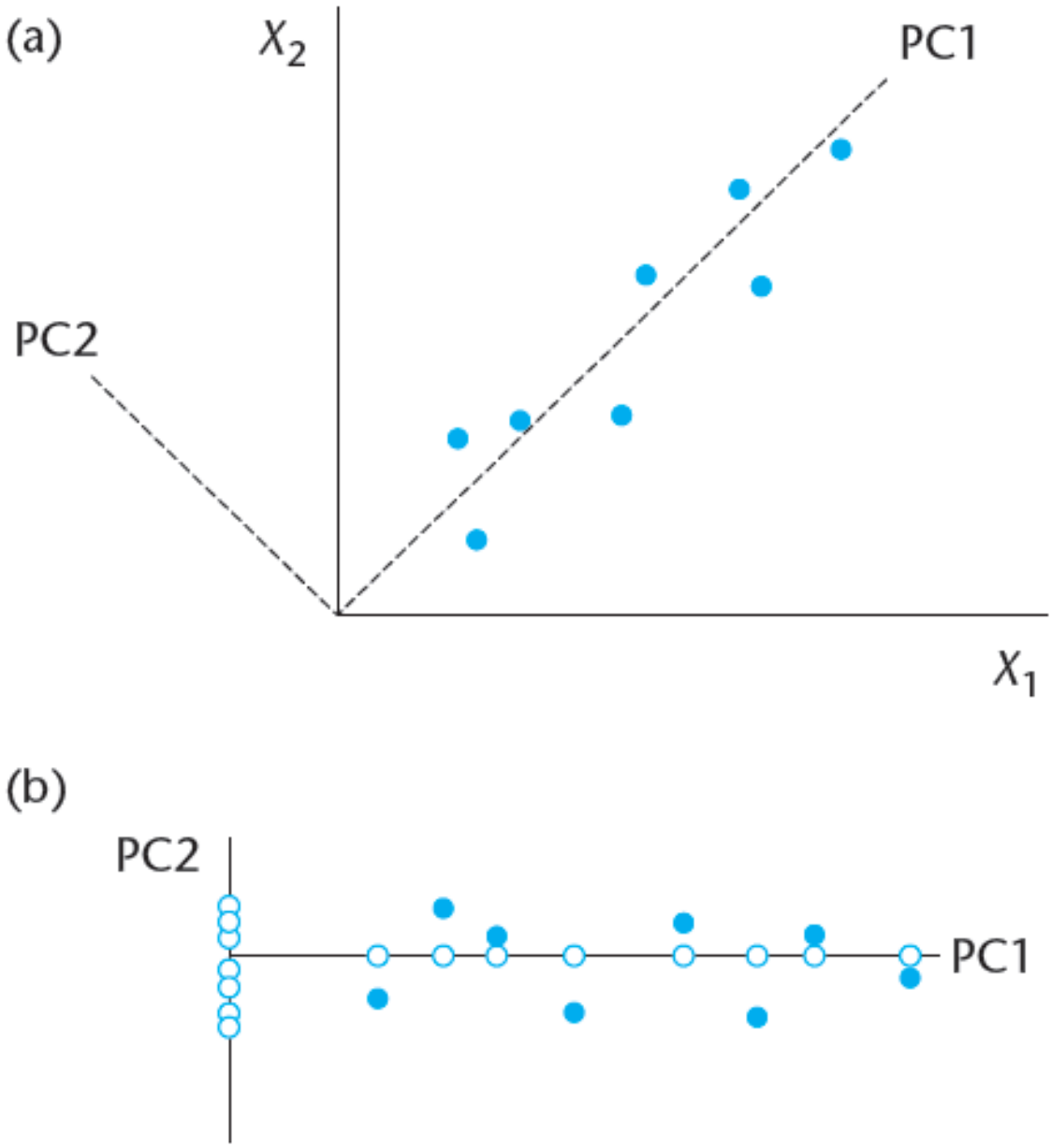


Figure 2-10 Diagram showing principal components 1 and 2 in a situation with only two variables. (a) The two principal components, PC1 and PC2, for the variables X1 and X2. (b) Datapoints referred to their principal components axes. Blue circles represent datapoints, white circles their projection onto the axes. Figure reproduced from Miller 2010.

Linear discriminant functions aim to maximise the separation between groups, whilst minimising the within group variance. Each function represents a line that describes maximal variation between two groups: new datapoints can then be

classified based on their position along this line relative to the known datapoints from each of the two groups. This can be extended to discriminate multiple groups. The difference between PCA and LDA is summarised in Figure 2-11.

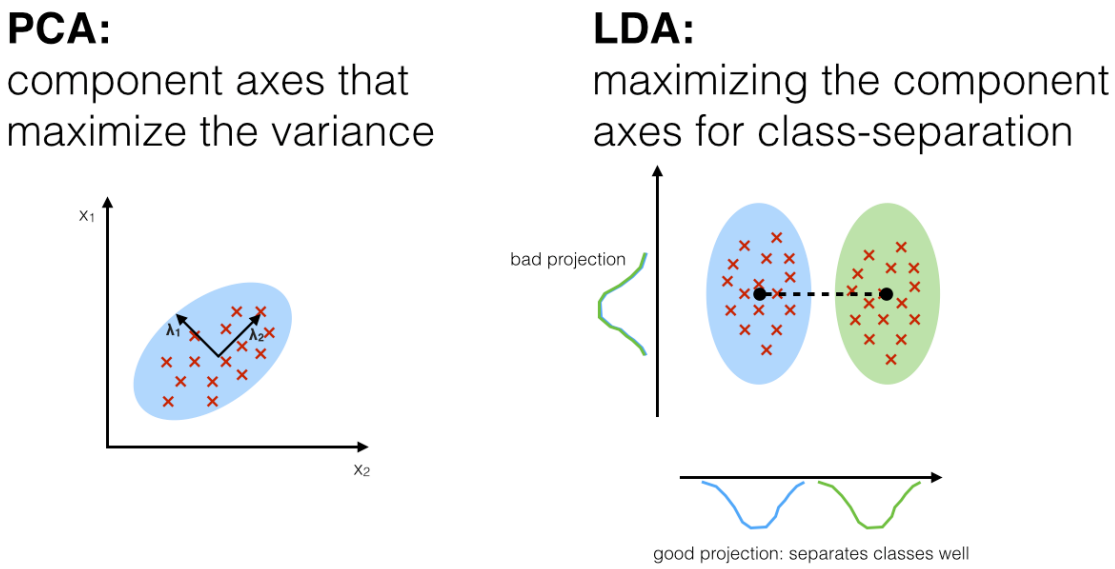


Figure 2-11 Comparison of PCA and LDA. Reproduced from: http://sebastianraschka.com/Articles/2014_python_lda.html (accessed 05/10/15).

This can then be used to construct a classification model onto which new 'test' data can be projected, with the model predicting which groups the data belong to.

Depending upon where the decision boundary is drawn to classify one group versus another, the sensitivity and specificity of the classification model can be adjusted. For example, for a screening test where high sensitivity is an important feature, classification which favours the 'positive' or high-risk group may be preferred, giving a high sensitivity at the expense of a low specificity. In Figure 2-12 for example, the linear decision boundary could be moved 'lower'

as plotted, to ensure that no red dots are misclassified (i.e. high sensitivity), at the expense of poor specificity.

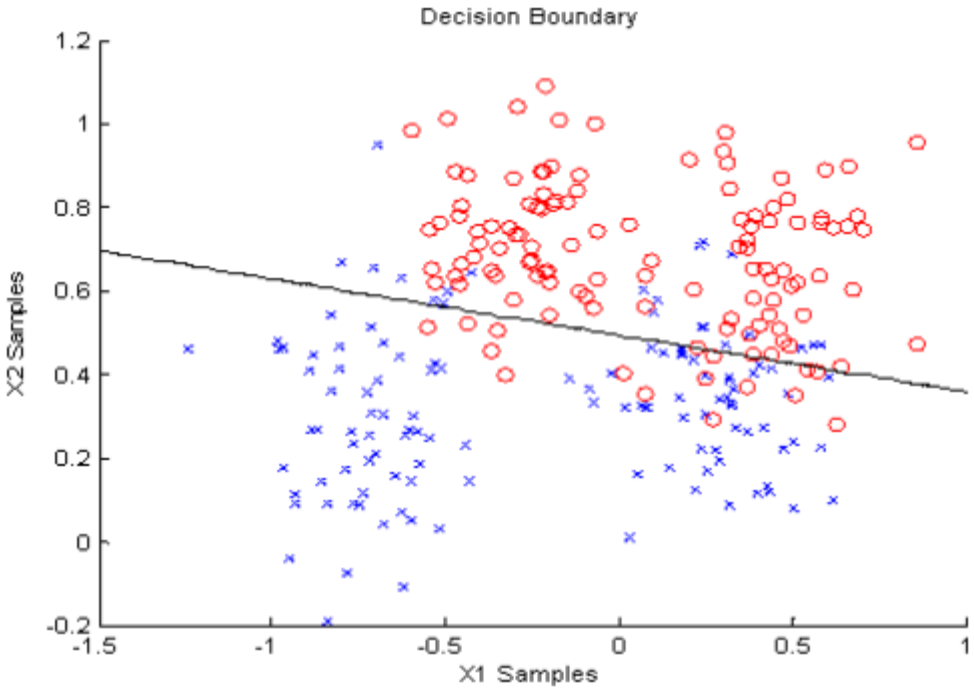


Figure 2-12 A linear decision boundary in LDA. Reproduced from: http://www.ece.umn.edu/users/cherkass/ee4389/FDA.html (accessed 05/10/15).

The performance of the classification model can be tested using the training dataset, however this risks providing a falsely optimistic measure of performance because the data being projected onto the model is contained within the training dataset itself. This 'overfitting' can be avoided by using separate training and test datasets, whereby none of the test data forms part of the training data used to develop the model. However, it is desirable to maximise the data included in the training set (particularly when tissue samples are scarce) to capture as much of the normal variation within groups as possible and build an accurate model. Consequently other techniques can be used to validate performance that achieve this without compromising validity by including training data in the test model.

Leave one sample out cross validation (LOSOCV) tests the performance of the training model by removing the sample being tested from the training dataset. Thus the sample in question cannot influence the model and result in overfitting, but n-1 samples can be included to give an optimal training dataset. This is repeated for every sample in turn, so no sample is ever in both the training and test datasets at the same time.

2.4.4. Summary of key steps involved in applying spectroscopy as a diagnostic tool

1. Creation of spectral library of biological samples (tissue/cells etc.) showing a range of normal and pathological states, correlated with histopathological analysis of each sample.

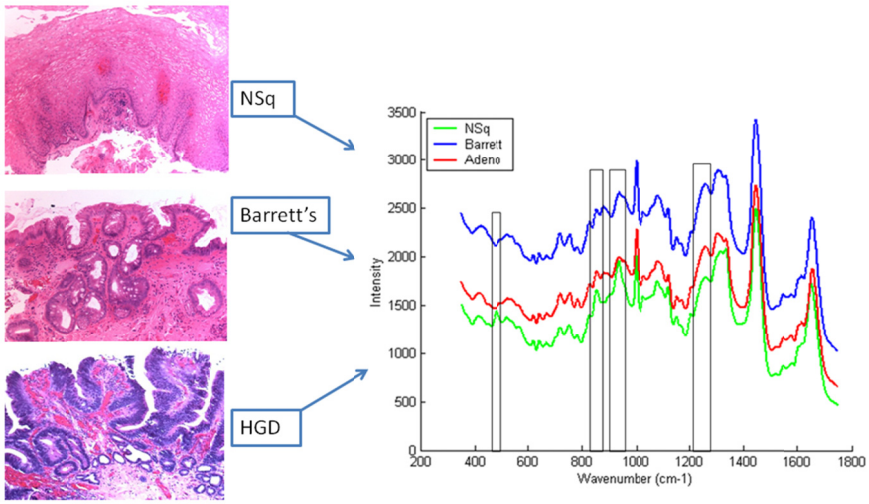


Figure 2-13 Mean Raman spectra from the three oesophageal pathology groups are shown in the right hand plot. Images of histological stained sections from oesophageal biopsies, showing normal squamous epithelium (NSq), Barrett's oesophagus and high grade dysplasia (HGD) are shown on the left (reproduced from Old et al. 2014).

2. Chemometric modelling to formulate algorithms that can be applied to interpret new samples.

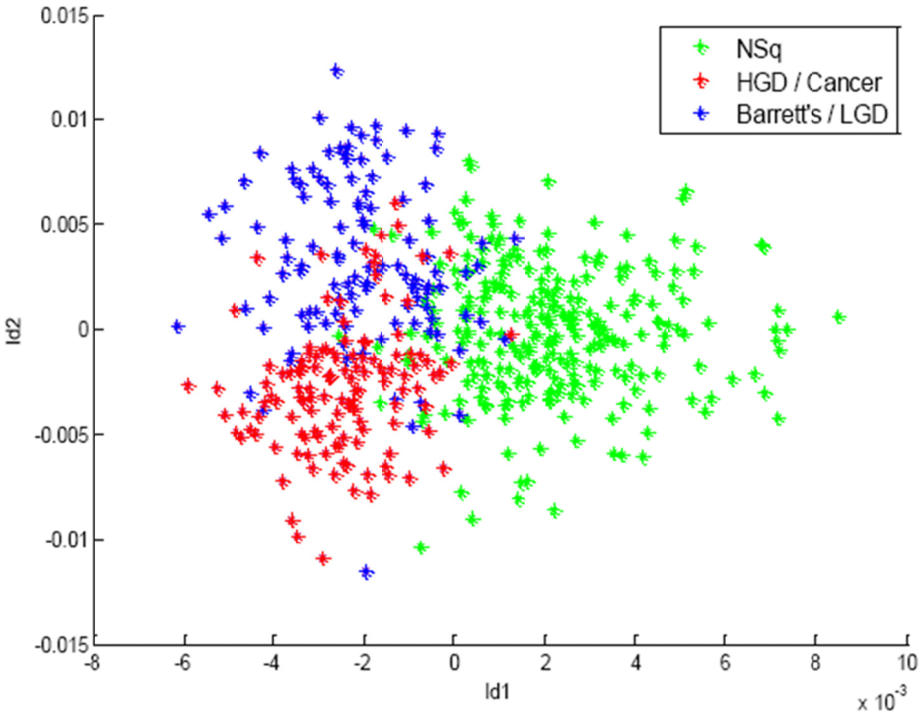


Figure 2-14 PCA-fed LDA modelling of oesophageal tissue samples measured with Raman probe, classified using 3-group model as Normal Squamous (NSq), high grade dysplasia (HGD)/cancer or Barrett's oesophagus/low grade dysplasia (LGD) (reproduced from Old et al. 2014).

3. Development of products for clinical use, such as new devices for *in vivo* diagnostics or new benchtop tools for automated histopathology. This stage may involve regulatory approval and clinical trials.

2.5. Summary of potential diagnostic applications of vibrational spectroscopy in the oesophagus

To date, much work has been undertaken in applying these spectroscopic methods to discriminate between disease states across a wide range of pathologies and organ systems, but as yet none have entered routine clinical practice. Current histopathological analysis of tissue relies largely on changes in tissue structure to classify disease processes: it is possible that the biochemical changes detected with spectroscopy may precede such changes, thus paving the way for earlier detection of disease, and facilitating rapid treatment.

Cancer diagnostics uses many different imaging modalities depending on the disease and presentation of the patient, but remains heavily reliant on tissue diagnosis by histopathologists to confirm malignancy. This is necessarily invasive, requiring removal of a biopsy sample of tissue, and usually requires tissue preparation involving fixation, sectioning and staining. These preparation processes and analysis by a skilled pathologist incur costs, and add a delay to the diagnosis of several days from the point of tissue sampling. Histopathologists analyse largely morphological information from the tissue specimens (sometimes with the additional benefit of immunohistochemical stains for specific molecules), and subjectivity in interpretation of these features is a recognised difficulty in certain conditions (Bennett et al. 2012; Montgomery et al. 2001).

Thus, novel methods to detect cancer in its early stages, or in a pre-malignant phase before it becomes an invasive tumour, are highly sought after, and the potential of vibrational spectroscopy to achieve this has been recognised.

Recent developments in spectroscopic instrumentation and spectral processing have advanced these technologies to a point where this potential is on the verge of translation into clinical reality. The development of fibre-optic probes has enabled access to a variety of epithelial lined body cavities, whilst mapping techniques allow interrogation of *ex vivo* samples, and 'Deep Raman' is being developed for non-invasive diagnosis in solid organs. The faster throughput of infrared spectroscopy lends it to rapid analysis of samples in the laboratory. Thus possible applications include automated screening of tissue biopsy processing in Barrett's surveillance, or analysis of oesophageal cell samples using newer methods of cell collection.

SECTION B: FTIR SPECTROSCOPY FOR DIAGNOSIS OF BARRETT'S OESOPHAGUS AND ASSOCIATED NEOPLASIA

CHAPTER 3: INTRODUCTION AND OBJECTIVES

3. Introduction and objectives

The current BSG guidelines list several future developments that would 'revolutionise the care of individuals with Barrett's oesophagus and should be priorities for policy makers and funders', of which the number one item listed is 'a non-endoscopic test(s) for diagnosis and surveillance' (Fitzgerald et al. 2014). Development of an accurate, minimally invasive, relatively low-cost test could radically alter current models of endoscopic surveillance, and overcomes the greatest obstacles to screening for Barrett's, namely the costs and acceptability/risks of endoscopy.

Swallowed cell collection devices offer a potential non-endoscopic means of sampling from the oesophagus. Balloon collection devices have been used in eastern Asia for several decades as a screening tool for squamous cell carcinoma of the oesophagus (Dawsey et al. 1997; Shu 1983; Spechler 1997). A balloon device has been trialled for Barrett's-associated neoplasia, but challenges included inadequate cell collection and reduced sensitivity for dysplasia (Falk et al. 1997). Results from the same study suggested that brush cytology has the potential to collect a representative sample, but the difficulty of interpretation of low grade dysplasia may limit the sensitivity using conventional cytological assessment.

If a swallowed cell collection device were to be used as a screening tool for Barrett's this would pose a number of challenges for conventional cytological assessment. Firstly, oesophageal cytology is performed relatively infrequently,

and expertise in this field is correspondingly limited. Secondly, assessment of cells may be challenging and shows variable correlation with histology taken contemporaneously (Geisinger et al. 1992; Hardwick et al. 1997; Hughes et al. 1998). Thirdly, a swallowed device that enters the stomach may have glandular cells from the stomach which must be differentiated from oesophageal glandular metaplasia. Fourthly, if expert cytological assessment were required for every sample if applied as a population screening tool, this would require a great deal of resources.

The recently developed Cytosponge™ device aims to overcome a number of these challenges through immunostaining with TFF3 (as explained in the literature review, section 1.8). A large case-control study in the UK is currently underway to evaluate this device as a means of detecting Barrett's oesophagus.[2]

Infrared spectroscopy has been used to discriminate accurately between pathology groups in a range of pathologies and organ systems, including between cells in the cervix, head and neck ('spectral cytopathology'). To date there have been a small number of studies investigating IR in the oesophagus. These studies have attempted to characterise the spectra of normal squamous, Barrett's, and adenocarcinoma tissue samples using FTIR mapping in formalin-fixed, paraffin embedded tissue (Quaroni and Casson 2009), or ATR in fresh-frozen sections to discriminate and classify normal squamous, Barrett's, dysplastic and gastric mucosa (Wang et al. 2007). Another study used FTIR mapping to produce spectral maps from 2 patients with oesophageal adenocarcinoma based on peaks at 1234cm^{-1} (representing the phosphodiester

[2] This trial has now recently reported and the published findings are discussed in the Summative Discussion.

peak, which they used as an indicator of nuclear content), and 1650cm^{-1} (the amide I peak, which they used as an indicator of cytoplasmic content) (Amrania et al. 2012).

A further study used an ATR probe in fresh frozen sections to discriminate normal squamous from squamous cell carcinoma (Wang et al. 2003).

As yet, no studies have investigated the use of FTIR mapping to classify oesophageal cells. The application of these techniques to cell samples taken using a cell collection device in the oesophagus could potentially show highly accurate discrimination between pathology groups, and provide an automated means of analysis. The ability to differentiate normal squamous cells from abnormal cells (Barrett's metaplasia or Barrett's-associated neoplasia) among screened patients would identify those who would need endoscopy. Accurate discrimination between Barrett's metaplasia and Barrett's-associated neoplasia could potentially be used among those with Barrett's oesophagus in place of current endoscopic surveillance.

3.1. Objectives

1. Investigate the ability of FTIR mapping to classify oesophageal pathology using oesophageal tissue samples and elucidate key spectral differences between pathology subtypes.
2. Investigate the ability of infrared spectroscopy to classify Barrett's and Barrett's neoplasia using oesophageal cell samples.

a. Evaluate the ability of infrared spectroscopy to discriminate between normal squamous, non-dysplastic Barrett's and Barrett's-associated neoplasia in oesophageal cell samples.
b. Evaluate the ability of infrared spectroscopy to discriminate between Barrett's metaplasia and glandular cells collected from the stomach.

CHAPTER 4: FTIR MAPPING OF OESOPHAGEAL TISSUE

4. FTIR mapping of oesophageal tissue

4.1. Ethical approval and informed consent

All patients participating in the study provided informed consent. Each patient received a standardised information leaflet explaining the study (Appendix V). Patients had time to read the leaflet prior to a face-to-face discussion with the researcher, which offered an opportunity to ask any questions about the study. If happy to proceed, patients then provided their consent to participate, and signed a standardised consent form (Appendix IV). No samples were taken from patients unless they had provided informed consent.

Ethical approval for the study was in place at the start of the project, as this project formed part of the ongoing work in the Biophotonics Research Unit programme for improving diagnosis in oesophageal malignancy. Ethical approval had previously been obtained from the Gloucestershire Local Research Ethics Committee (Appendix VI).

4.2. Sample collection

Samples were collected from patients undergoing scheduled endoscopy, either as part of Barrett's surveillance or indicated for other clinical grounds (e.g. previous finding of dysplasia) or from patients undergoing surgery for oesophageal cancer.

Since serial tissue sections were required to allow accurate correlation with histology (discussed further in section 4.4 below), only endoscopic resection or

samples taken at oesophageal resection were used in this study. Endoscopic resection was performed using diathermy, and a portion of the section sent for immediate histology with the remainder taken for research. The research sample was then placed into a cryovial (Thermo Scientific, Waltham, MA, USA) and immediately snap frozen in liquid nitrogen, before storage at -80°C.

For the samples collected at surgery for oesophageal cancer, once the oesophagus was removed from the patient it was opened longitudinally, washed in distilled water, and samples were collected using a scalpel. Samples were placed in cryovials and snap frozen as above.

4.3. Sample preparation

Fresh frozen samples were cut into thin tissue sections for measurement and analysis. 8μm thick sections were cut and placed on calcium fluoride slides (CaF_2) for FTIR measurement, with a contiguous tissue section being taken and placed on a glass slide for haematoxylin and eosin (H&E) staining, followed by conventional histological assessment.

The use of glass slides for infrared analysis is limited because infrared does not penetrate glass in the fingerprint region, hence calcium fluoride slides are preferred: calcium fluoride has no absorption peaks in the measured wavenumber range.

The tissue section placed on calcium fluoride was then stored at -80°C until measurement.

4.4. Histology correlation: identifying regions to be measured

Taking contiguous sections ensures that the tissue section used for analysis will very closely match the histology of the section used for H&E staining. This allows accurate correlation of the histology findings with the spectroscopic data. Tissue sections were reviewed by a Consultant Pathologist with a special interest in gastrointestinal pathology. For this study, in addition to classifying the pathology of each tissue section, areas within that section were highlighted by the pathologist to identify areas of homogeneous pathology e.g. dysplastic epithelium. Epithelial regions of interest were identified containing homogeneous pathology for measurement. This allowed targeted FTIR mapping of the contiguous tissue section, and ensured that, as far as possible, only a single tissue pathology was contained in the regions measured.

Subsequently, tissue sections were reviewed independently by a second Consultant Pathologist with a special interest in gastrointestinal pathology, in accordance with BSG guidelines on reporting of dysplasia. Appendix VII (i) shows the classification of each tissue sample by both pathologists. This showed agreement in all normal squamous, Barrett's and adenocarcinoma samples. There was discrepancy over 4 of the 5 dysplasia samples, with disagreement over classification of HGD versus LGD, but all samples were classified as dysplastic by both pathologists. The overall agreement was therefore 81.8% if the discordance over grade of dysplasia is included, or 100% if using a more simple classification of 'any dysplasia'. This high level of agreement between the pathologists is likely to be due to the selection of 'homogeneous' regions by the first pathologist. The effects of variation in

pathology classification on our results are discussed further in the discussion (section 4.11).

To identify the region to be measured, the unstained tissue section was examined under the microscope, and this image used to identify features of the tissue that correspond to regions on the H&E stained section (see Figure 4-1). Once the region of interest had been identified on the calcium fluoride section, an area within that region of interest was chosen to be measured.

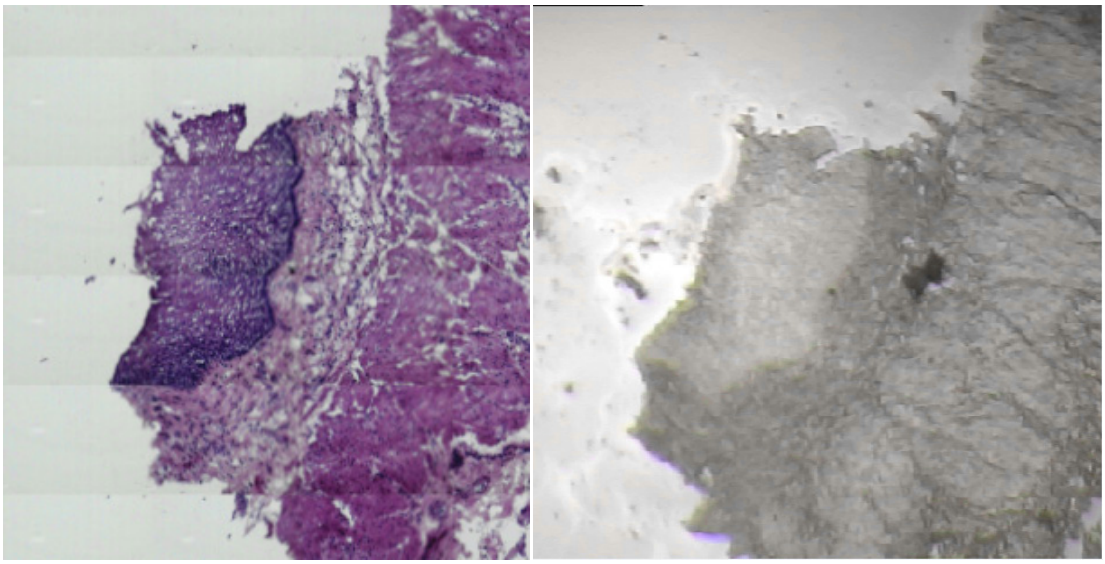


Figure 4-1 Correlating H&E stained section with contiguous unstained tissue section (small region of normal squamous epithelium, area highlighted further in Figure 4-2 below).

4.5. Sample measurement

FTIR measurements were performed in the Department of Physics at the University of Exeter, using an Agilent® 620 FTIR microscope connected to an Agilent® 670 spectrometer.

The Agilent® system has a 128x128 pixel focal plane array detector and uses a high energy Globar mid-infrared light source to simultaneously acquire thousands of spectra. A schematic diagram of an FTIR spectrometer is shown in Figure 2-9 in the literature review.

On the Agilent® system there is not an enclosed measurement chamber, but to minimise the presence of atmospheric water vapour the sample is continuously purged with anhydrous air. Prior to sample measurement the liquid nitrogen reservoir was filled with liquid nitrogen to ensure adequate cooling of the focal plane array detector.

The tissue section on calcium fluoride was defrosted prior to measurement. Once the sample slide was mounted on the FTIR microscope, system calibration was performed by choosing a blank region of the slide, and adjusting the integration time and condenser objective to optimise the intensity of the IR beam on the sample.

A background spectrum was measured on a blank region of the sample slide prior to measurement of each sample, to allow correction for absorbance from the atmosphere, the system and the slide. Background measurement was taken using 256 scans at $4cm^{-1}$ resolution for wavenumbers 800-$4000cm^{-1}$.

Regions of homogeneous pathology were identified as described above, and the region of interest was identified on the slide using the microscope. FTIR measurements were performed using the focal plane array detector of the Agilent 620 microscope in transmission mode, using ‘High-magnification’ mode which gives a pixel size of 1.1µm, with a spectral resolution of $4cm^{-1}$. The focal plane array allows measurement of 128x128 pixels in a single shot, so with a

1.1μm pixel size this gives an approximately 140μm x 140μm region of measurement for a single 'tile'.

Consequently, 2 separate 140μm x 140μm regions were selected within the region of interest to be measured. This size corresponds to one 'tile' of the Agilent detector, allows rapid acquisition (approximately 12 minutes per tile using 64 scans per pixel), and provides a large number of spectra for analysis (16,384 per tile), whilst being a sufficiently small region to be confident of measuring homogenous tissue pathology.

4.6. Data pre-processing

The spectral maps collected are stored as a hyperspectral image or hypercube. This can be considered as a 2-dimensional map with co-ordinates in the X and Y plane, with the 3rd dimension representing the complete spectral information for a given pixel. An example of an unprocessed spectral map with its corresponding white light image is shown in Figure 4-2 below. The unprocessed map simply plots the maximum intensity of the signal at each pixel.

The raw spectrum measures the amount of light transmitted through the tissue, (transmittance) relative to a background reading taken prior to measurement: this is converted to the amount of light absorbed by the tissue using the Beer-Lambert Law as follows:

$$A = -\log_{10}(T)$$

Where A= Absorbance, T= Transmittance

The Agilent system does this calculation automatically and the collected data is then stored as absorbance rather than transmittance spectra.

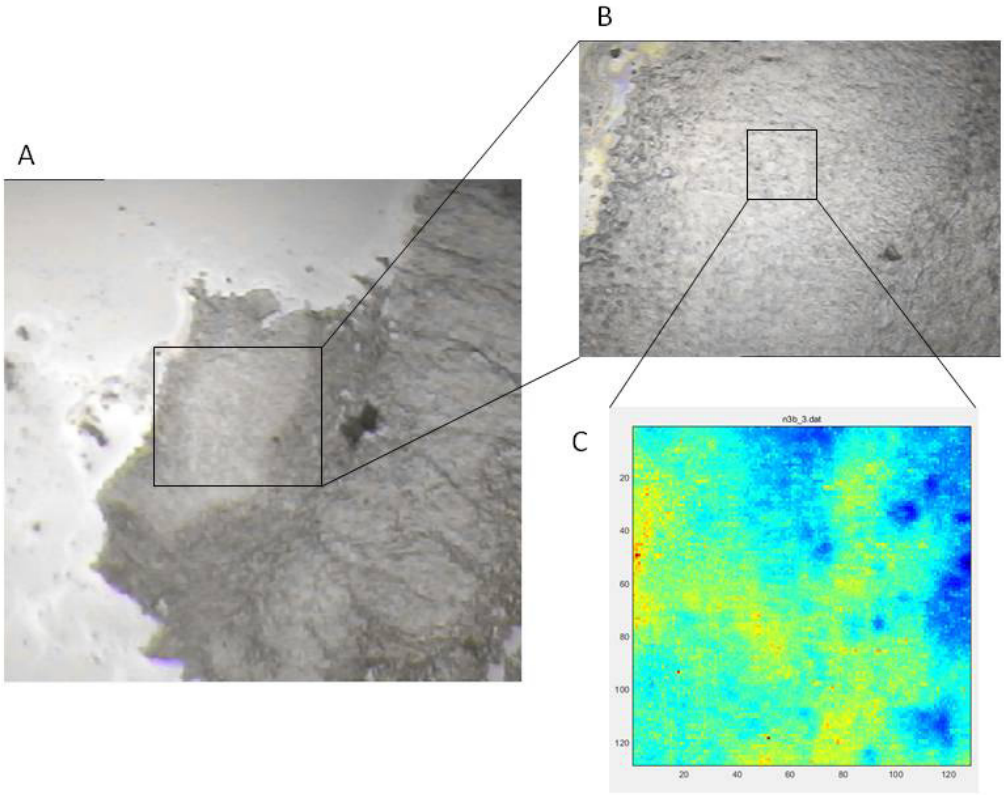


Figure 4-2 White light image and unprocessed spectral map (sample N3), showing (A) region identification using 4x objective, (B) selection of region of interest to map using 15x objective, and (C) unprocessed FTIR spectral map.

The measured wavenumber range is 800-4000cm^{-1}. In this large dataset, to focus on the most discriminatory 'fingerprint' region and allow more rapid analysis, the wavenumber range was truncated to 950-1800cm^{-1}. This risks losing some information contained in the higher wavenumber range, but was felt to be an acceptable trade off based on the key discriminatory wavenumber regions identified in previous work within the group.

To minimise differences in absolute signal intensity due to differences in tissue thickness, or system effects such as source or detector variations, the dataset was vector normalised using the Standard Normal Variate normalisation method. Vector normalisation is a well recognised pre-processing technique for reducing scaling effects in infrared microspectroscopy (Diem et al. 2004).

The dataset is then mean centred – this is useful prior to principal component analysis as it ensures the principal components focus on variance from the mean rather than variance from the origin, giving more meaningful information about variance between spectra.

4.7. Data analysis

The principles underlying the techniques chosen for data analysis are described in greater detail in section 2.4 above.

Principal component analysis (PCA) was used as a data reduction technique to identify variance between the spectra. The first 50 principal components (PCs) were identified: this aimed to optimise classification performance without 'over-fitting' data using random noise. Analysis of variance (ANOVA) was then performed to identify which of those PCs provide the greatest discrimination between pathology groups, with a confidence threshold of 0.99 for inclusion in the training model.

Linear discriminant analysis (LDA) was then performed using those PCs selected by ANOVA, with the samples assigned into groups according to consensus pathology reporting, to build diagnostic classification models.

A variety of diagnostic classification models were then built, as described below. Each of these was tested using leave one sample out cross-validation (LOSOCV), with each spectrum being projected onto a training model using data from every sample except the one being tested (i.e. n-1 samples). This was repeated for every spectrum in the dataset. The predicted classification could then be used to generate a confusion matrix showing the true and false predictions for each spectrum, and allowed sensitivity and specificity of the technique to be calculated for individual spectra.

Once predictions had been calculated for individual spectra, it was possible to combine these results to generate predictions for entire sample maps. A 'threshold voting' system was developed which incorporates the prediction for each pixel of a map and combines this data to produce a classification for the whole sample. This is explained further below.

4.8. Summary of measurements

A total of 737,280 spectra were measured from 22 tissue samples, collected from 19 patients (some patients contributed tissue of more than one pathology type, for example if an area of normal squamous tissue and an area of adenocarcinoma were collected from the same patient). 140μm x 140μm FTIR maps were measured over 2 regions for each tissue sample (and there was one 'extra' sample region measured for one of the adenocarcinoma samples). The number of maps recorded and spectra measured for each pathology are shown in Table 4-1 below.

Table 4-1 Summary of measured samples

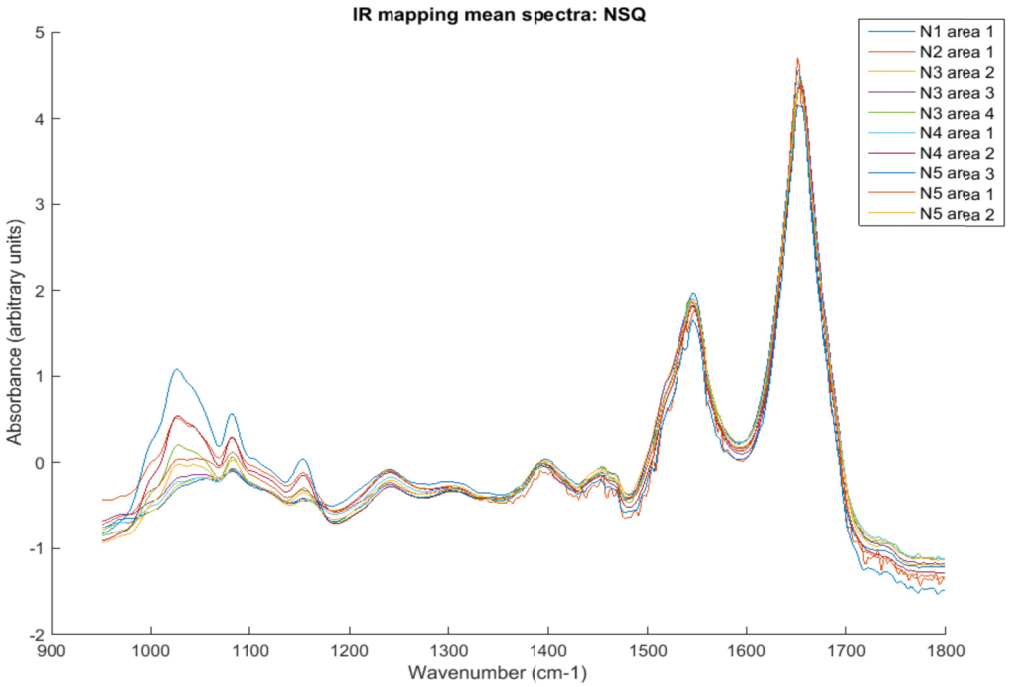
	No. of tissue samples	Total no. of FTIR maps measured	Total no. of spectra measured
NSQ	5	10	163,840
BE	6	12	196,608
Dysplasia	5	10	163,840
Adenoca.	6	13	212,992
TOTAL	22	45	737,280

NSQ normal squamous; BE Barrett's oesophagus; Adenoca. adenocarcinoma.

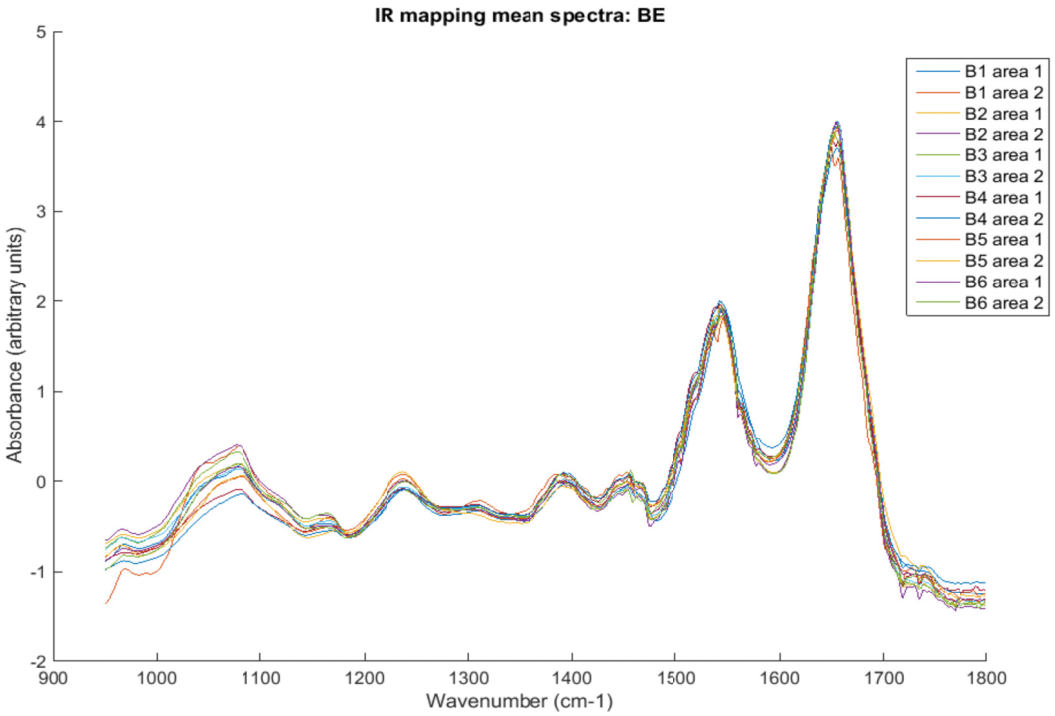
After pre-processing, the mean spectrum for each sample was reviewed to ensure all samples were valid measurements prior to inclusion as training data for classification models. The mean spectrum of each sample is presented in Figure 4-3 below.

The mean spectra from every sample were consistent with tissue spectra, and every sample was included for developing classification models. There was significant variance seen within some groups e.g. normal squamous around 1020cm^{-1}, and dysplasia around 1650cm^{-1}. Further analysis of the composition of these spectra, and differences between the pathology groups, is discussed in section 4.10 below.

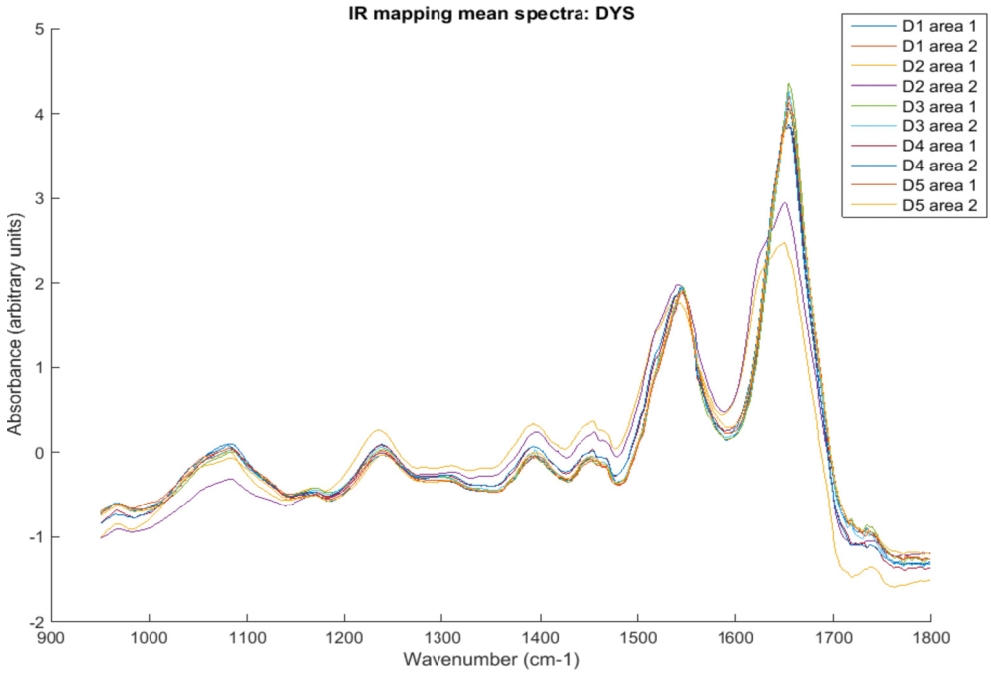
A


IR mapping mean spectra: NSQ
N1 area 1
N2 area 1
N3 area 2
N3 area 3
N3 area 4
N4 area 1
N4 area 2
N5 area 3
N5 area 1
N5 area 2
Absorbance (arbitrary units)
Wavenumber (cm-1)
5
4
3
2
1
0
-1
-2
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

B


IR mapping mean spectra: BE
B1 area 1
B1 area 2
B2 area 1
B2 area 2
B3 area 1
B3 area 2
B4 area 1
B4 area 2
B5 area 1
B5 area 2
B6 area 1
B6 area 2
Absorbance (arbitrary units)
Wavenumber (cm-1)
5
4
3
2
1
0
-1
-2
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

C



D

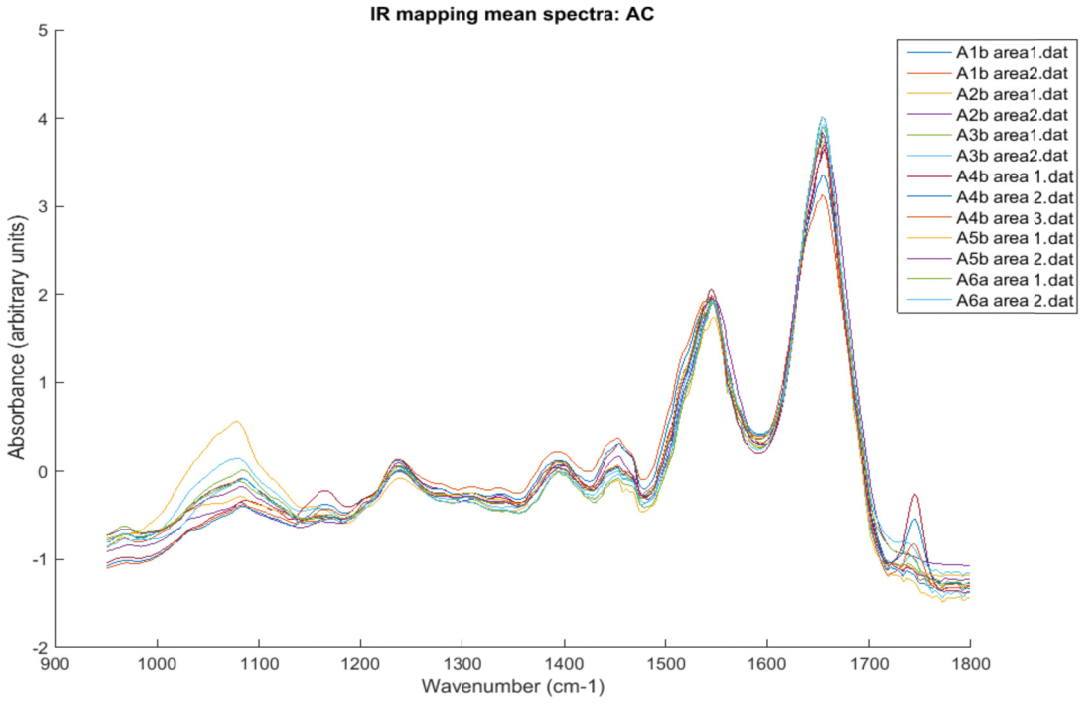


Figure 4-3 Mean spectra of IR mapping samples, presented per pathology group with (A) showing normal squamous (NSQ), (B) Barrett's oesophagus (BE), (C) dysplasia (DYS) and (D) adenocarcinoma (AC).

4.9. Diagnostic classification models

4.9.1. 3 group model

For the initial analysis the low grade dysplasia, high grade dysplasia and adenocarcinoma groups were combined for classification purposes, thus giving a '3 group model' with classification as either 'Normal squamous', 'Barrett's', or 'Dysplasia/adenocarcinoma'. This is for two reasons: firstly if FTIR mapping were to be used in practice for Barrett's surveillance, this classification would highlight the 'high risk' samples which would benefit from intervention. Current guidelines advise dual reporting of all samples containing dysplasia or adenocarcinoma, so one potential role for FTIR would be to act as the 'first pathologist' in screening samples to identify those high risk specimens requiring a second review.

Secondly, this project was intended to act as a preliminary study for cytology classification: cytology classification is more limited than histology and can only discriminate between normal, glandular (i.e. consistent with Barrett's) and dysplastic cells (for either dysplasia or adenocarcinoma). Consequently the biochemical differences between these groupings would be most useful in demonstrating feasibility to inform the cytology study.

PCA was used to identify the sources of variance within the dataset, with up to 50 principal components (PCs) chosen by ANOVA. The contribution of each of those PCs in describing the variance within the dataset is shown by the F ratio plot in Figure 4-4.

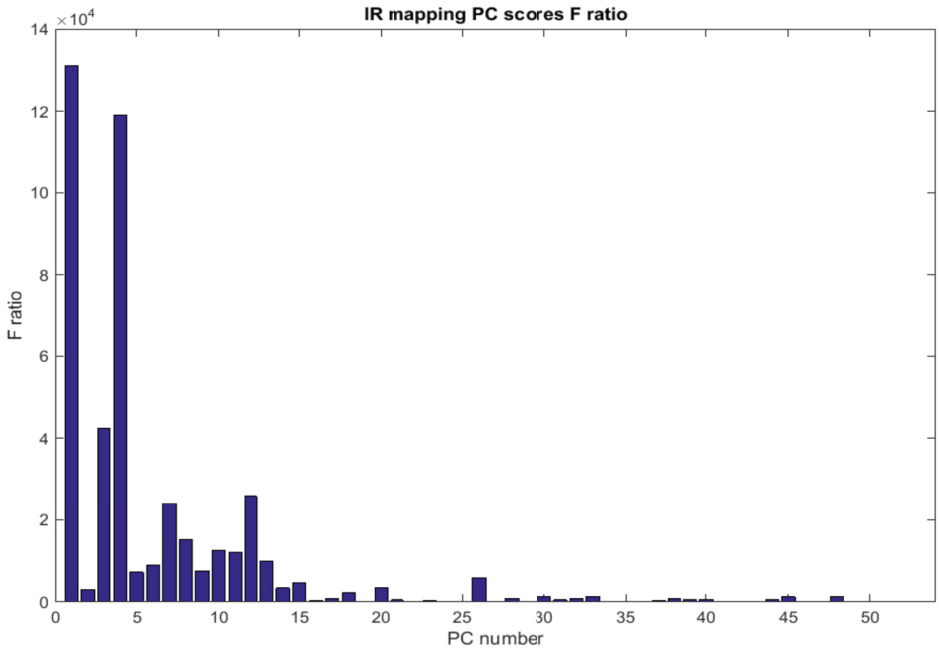


Figure 4-4 F ratio for each principal component (PC). The F ratio gives a measure of how much variation each PC contributes to the dataset.

The F ratio plot in Figure 4-4 shows that nearly all of the variance is described by the first 15 PCs, and there is very little contribution after this, suggesting the higher PCs are mostly random noise.

Figure 4-5 is a 3D plot showing the 3 most discriminatory PCs. This shows that the normal squamous and Barrett's spectra are fairly tightly grouped, but the dysplasia/adenocarcinoma spectra are far more heterogeneous. As a result, these 3 PCs separate the normal squamous spectra from the other 2 groups reasonably well, but there is considerable overlap between the Barrett's and dysplasia/adenocarcinoma spectra.

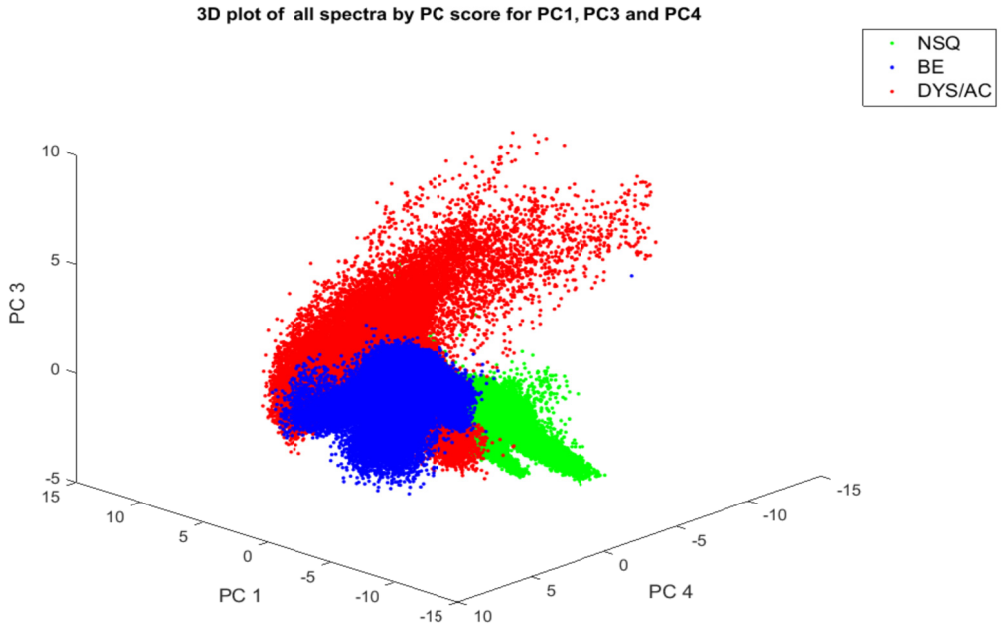


Figure 4-5 Scatter plot of all spectra, plotted by principal component (PC) score for 3 most discriminatory PCs: PC1, PC3 and PC4. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

LDA was used to identify differences between the groups and classify the spectra into pathology groups. The most discriminatory linear discriminant functions are plotted in Figure 4-6. Here, spectra from each of the 3 pathology groups are fairly tightly grouped. The normal squamous spectra are clearly distinct from the other 2 groups, but overlap between the Barrett's and dysplasia/adenocarcinoma groups persists.

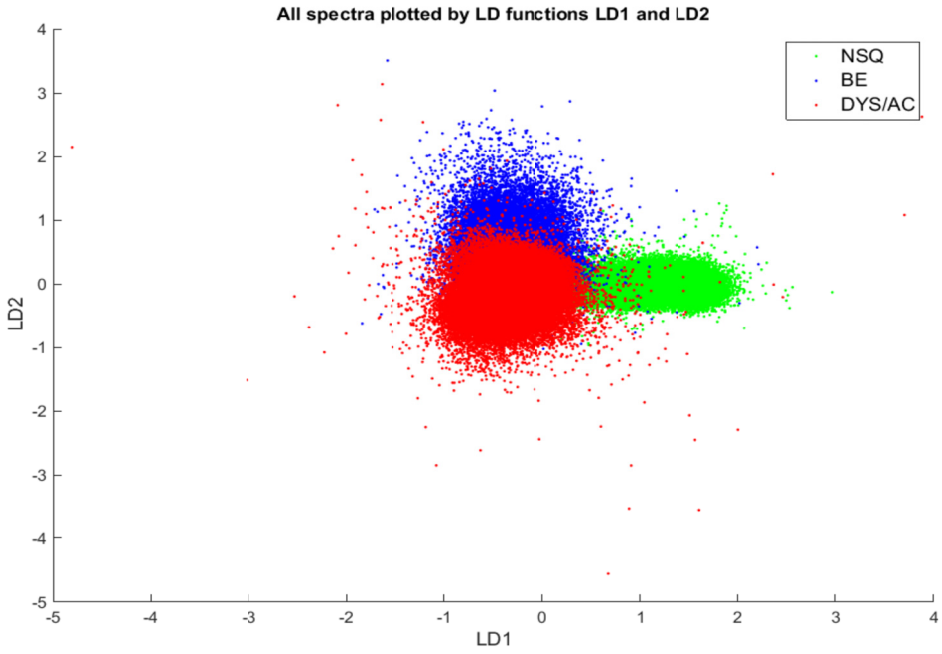


Figure 4-6 Scatter plot of all spectra, plotted by linear discriminant (LD) function for 2 most discriminatory LDs. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

Once PCA-fed linear discriminant analysis had been used to classify the spectra, the results were tested using leave one sample out cross validation. The results are shown in Table 4-2 below.

The classification model showed very good classification of normal squamous samples, with high sensitivity (93.6%) and very high specificity (99.9%): only 0.1% of spectra from tissue samples with a true underlying pathology (Barrett's/dysplasia/adenocarcinoma) were misclassified as normal by the model.

The classification of spectra from Barrett's and the dysplasia/adenocarcinoma was reasonably good, but less accurate than the normal squamous. The confusion matrix shows there was some cross over between these groups, with 57,645 dysplasia/adenocarcinoma spectra (15.3%) misclassified as Barrett's,

and 45,500 Barrett's spectra (23.1%) misclassified as dysplasia/ adenocarcinoma.

Table 4-2 Performance of the 3 group model: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology		
Predicted spectral pathology	**NSQ (no. of spectra)**	**BE (no. of spectra)**	**DYS/AC (no. of spectra)**
NSQ	**153358**	221	583
BE	4680	**150887**	57645
DYS/AC	5802	45500	**318604**

B

	Tissue pathology		
Performance of 3 group model	**NSQ**	**BE**	**DYS/AC**
Sensitivity (%)	93.6	76.8	84.6
Specificity (%)	99.9	88.5	85.8

To further test the validity of these results, a further analysis was performed for each sample which excluded other samples taken from the same patient, i.e. a leave one patient out cross-validation (LOPOCV). This goes a step further than LOSOCV to ensure that any patient-specific differences (rather than pathology-specific differences) are not included in the training model. The results are shown in Table 4-3.

Table 4-3 Performance of the 3 group model after LOPOCV: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology		
Predicted spectral pathology	**NSQ (no. of spectra)**	**BE (no. of spectra)**	**DYS/AC (no. of spectra)**
NSQ	**148668**	340	593
BE	8710	**124191**	76117
DYS/AC	6462	72077	**300122**

B

	Tissue pathology		
Performance of 3 group model	**NSQ**	**BE**	**DYS/AC**
Sensitivity (%)	90.7	63.2	79.6
Specificity (%)	99.9	84.3	78.2

The results in Table 4-3 show a very slightly poorer performance for normal squamous and dysplasia/adenocarcinoma, but a larger drop in the Barrett's sensitivity from 76.8% to 63.2%. The confusion matrix indicates that many more Barrett's spectra are misclassified as dysplasia/adenocarcinoma, but very few as normal squamous. This suggests that between patient differences contributed a small amount to the training model overall, but that this may have been a greater source of discrimination between the Barrett's and dysplasia/adenocarcinoma groups than the others.

Ultimately the purpose of FTIR mapping would be to classify each sample rather than individual pixels within a given sample. Consequently a system was required to convert the individual pixel classification into a prediction for the whole sample. In a theoretical situation where pixel classification was 100% accurate, then it might be logical to classify a sample with even 1 pixel containing evidence of adenocarcinoma as cancer, or at least 'high risk' for cancer. However, it was clear from the pixel classification performance that this level of accuracy was unrealistic with 2500 pixels per sample, and greater flexibility was required.

Projecting the pixel classification back onto the maps shows the distribution of the prediction within the sample. Although sample regions had been carefully selected to show a single pathology, in some cases the spectral predictions were different across a sample, so did not all match the histology classification. Projections of these predictions onto the maps could show if a region within a map was classified differently as shown in Figure 4-7.

There is a strong analogy with the methodology used by pathologists here – when reviewing specimens if a tiny area of cancer is seen then the sample is classified as showing cancer (even though the majority of the sample may be benign). Similarly, if a sample shows a region of Barrett's mucosa and a region of normal squamous mucosa, the specimen (and therefore the patient) will be classified as having Barrett's oesophagus. Effectively the 'worst' pathology present is that used to classify the sample.

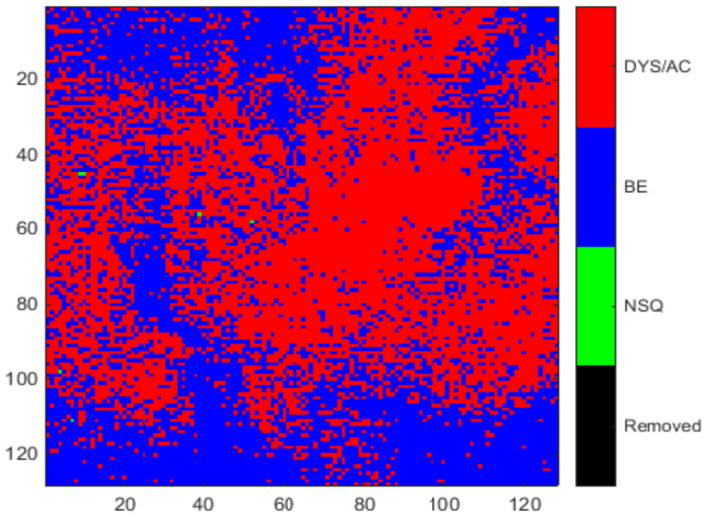


Figure 4-7 Projection of spectral predictions onto sample B4(2) (Barrett's oesophagus) showing regions of the map classified as dysplasia /adenocarcinoma.

We aimed to replicate this by applying a threshold voting system for each sample. After discussion with a pathologist, considering what would be appropriate to meet the aim of instituting FTIR as a sample screening technique (where high sensitivity is required), and taking into account what would be realistic to allow a degree of pixel misclassification, we decided that a threshold of 30% would be appropriate i.e. each sample would be classified as the 'worst' pathology present, provided >30% of the pixels in that sample were classified as that pathology. The results of the voting classification are shown in Figure 4-8 below, with the sensitivity and specificity using a 30% threshold in Table 4-4.

Using the 30% threshold, every normal sample was classified correctly. If used as a tool to screen samples to identify high risk cases for further pathology review, a high sensitivity would be required to ensure no samples were missed. The sensitivity for detecting dysplasia/adenocarcinoma samples was 95.7%, with one dysplastic sample being misclassified as Barrett's oesophagus. Using

this low threshold of 30%, 3 Barrett's samples were misclassified in the dysplasia/adenocarcinoma group.

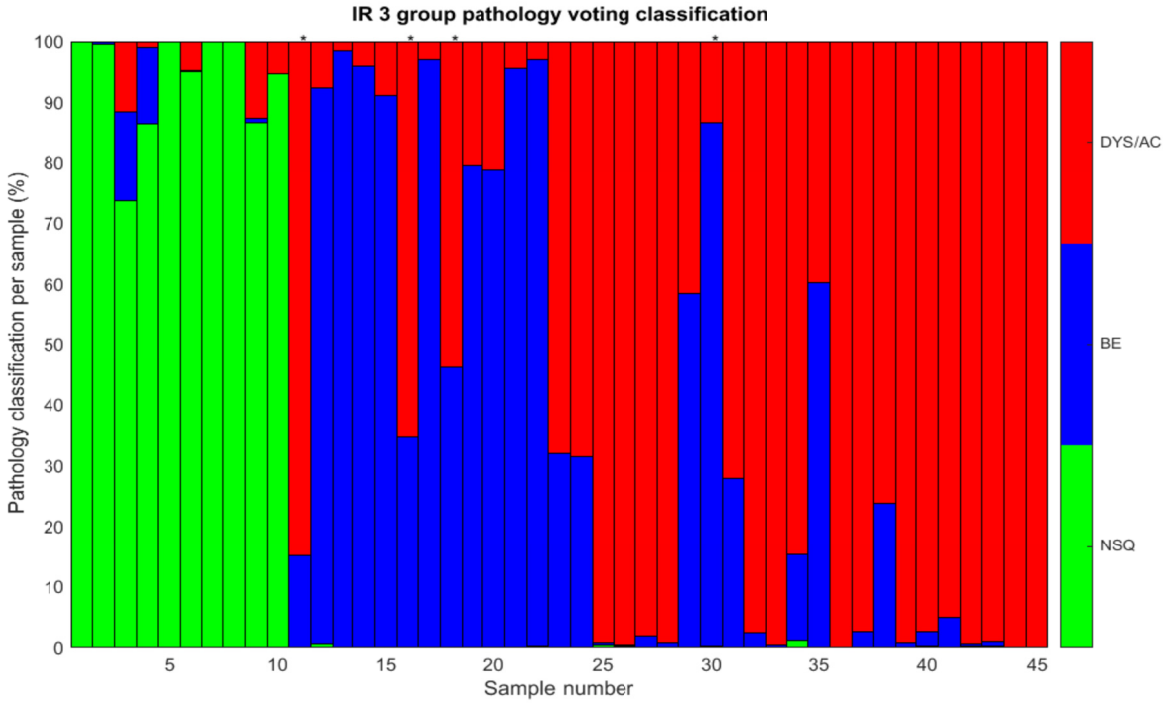


Figure 4-8 Classification of each sample using voting classification for FTIR mapping 3 group model. The proportion of pixels classified as each pathology is shown for each sample in turn. Samples misclassified using the 30% threshold are marked with an asterisk at the top of the bar. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma. Samples 1-10 are NSQ, 11-22 are BE, 23-45 are DYS/AC.

The voting classification in Figure 4-8 shows the spectra from normal squamous samples were uniformly predicted correctly. The misclassification of some spectra from Barrett's samples as dysplasia/adenocarcinoma (and vice versa) is seen, and for some samples this is sufficient to cause misclassification of the whole sample.

Table 4-4 Performance of the 30% threshold voting 3 group model: (A) confusion matrix showing classification of samples (with correctly classified samples in **bold**), and (B) sensitivity and specificity of the model for classifying samples. (C) List of samples misclassified by the model. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	True pathology classification		
Predicted	**NSQ**	**BE**	**DYS/AC**
NSQ	**10**	0	0
BE	0	**9**	1
DYS/AC	0	3	**22**

B

Sample pathology	Sensitivity	Specificity
NSQ	100	100
BE	75.0	96.7
DYS/AC	95.7	86.4

C

Sample no.	True pathology	Predicted pathology
B1(1)	BE	DYS/AC
B3(2)	BE	DYS/AC
B4(2)	BE	DYS/AC
D4(2)	DYS	BE

Looking at the 4 samples which were misclassified (see Table 4-4 C), it is noticeable that each of these samples was rather poorly preserved, as can happen with fresh frozen sections. This results in slight loss of tissue structure and the presence of small holes within the tissue, and it is possible that this may have contributed to the failure to classify these sections correctly. However,

there were 2 other tissue sections which were equally poorly preserved (B5 and D5) where the classification model uniformly predicted the correct pathology.

One other sample, A2(1), although classified correctly, showed heterogeneity in its spectral classification. This sample had a tear in the tissue section making correlation of the region of interest with H&E more challenging, and it is possible that imprecise selection of the measurement region contributed to this.

Table 4-5 Performance of the 30% threshold voting 3 group model using LOPOCV: (A) confusion matrix showing classification of samples (with correctly classified samples in **bold**), and (B) sensitivity and specificity of the model for classifying samples. (C) List of samples misclassified by the model. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	True pathology classification		
Predicted	**NSQ**	**BE**	**DYS/AC**
NSQ	**10**	0	0
BE	0	**7**	1
DYS/AC	0	5	**22**

B

Sample pathology	**Sensitivity**	**Specificity**
NSQ	100	100
BE	58.3	97.0
DYS/AC	95.7	77.3

C

Sample no.	**True pathology**	**Predicted pathology**
B1(1)	BE	DYS/AC
B3(2)	BE	DYS/AC
B4(2)	BE	DYS/AC
B5(1)	BE	DYS/AC
B5(2)	BE	DYS/AC
D4(2)	DYS	BE

The threshold voting system was then repeated using the LOPOCV results, shown in Table 4-5. As with the individual spectra results the sensitivity for Barrett's is lower than for LOSOCV, with several samples misclassified as dysplasia/adenocarcinoma, and the specificity of the dysplasia/adenocarcinoma group is correspondingly reduced.

An alternative approach to the threshold voting system is to use a very high threshold, and accept that a proportion of the samples will therefore not be classified and the model will return a result of 'unknown pathology'. This could be useful as a clinical application if sole diagnosis were to be performed by FTIR mapping, with only those samples labelled as unknown by FTIR then needing review.

Table 4-6 Performance of the 90% threshold voting 3 group model: (A) confusion matrix showing classification of samples (with correctly classified samples in **bold**), (B) sensitivity and specificity of the model for classifying samples (not including unclassified samples). NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	True pathology classification		
Predicted	**NSQ**	**BE**	**DYS/AC**
NSQ	**7**	0	0
BE	0	**7**	0
DYS/AC	0	0	**15**
UNKNOWN	3	5	8

B

Sample pathology	**Sensitivity**	**Specificity**
NSQ	100	100
BE	100	100
DYS/AC	100	100

A 90% threshold was chosen for this model, aiming to include only samples that the model was able to predict with high confidence, but allowing a small margin for error on pixel classification. Classification performance of the 90% threshold voting model is shown in Table 4-6. Using a 90% threshold gives 100% accurate diagnosis for those samples the model is able to classify, but results in 35.6% of samples being unclassified. This provides further evidence that the classification model, where confident in its predictions, is accurate, but leaves a high proportion of samples unclassified. This could nonetheless be useful in clinical practice as an automated system with 'knowns' fully trusted and unknowns sent for pathology review.

4.9.2. 2 group classification models

4.9.2.1. 'Normal' versus 'abnormal'

A 2 group model was also evaluated, combining the Barrett's, dysplasia and adenocarcinoma samples: this could be clinically applicable if used simply to classify samples as 'normal' versus 'abnormal', with abnormal samples selected for further review. The results of the 2 group model are shown in Table 4-7 and Figure 4-9 below.

The ability of the model to discriminate between normal and abnormal pathology is very good, with sensitivity and specificity for classification of individual spectra of 93.6% and 99.8% respectively. When translated into whole sample classification, the model achieves 100% accuracy.

Table 4-7 Performance of the 2 group model NSQ vs BE/DYS/AC. Results for individual spectra are shown in A) confusion matrix (with correctly classified spectra in **bold**), and (B) sensitivity and specificity. Threshold voting classification of whole samples is shown in C) confusion matrix and D) sensitivity and specificity. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology	
Predicted spectral pathology	**NSQ (no. of spectra)**	**BE/DYS/AC (no. of spectra)**
NSQ	**153349**	1304
BE/DYS/AC	10491	**572136**

B

	Tissue pathology	
Performance of 2 group model	**NSQ**	**BE/DYS/AC**
Sensitivity (%)	93.6	99.8
Specificity (%)	99.8	93.6

C

	Tissue pathology	
Predicted sample pathology	**NSQ**	**BE/DYS/AC**
NSQ	**10**	0
BE/DYS/AC	0	**35**

D

	Tissue pathology	
Performance of 2 group model	**NSQ**	**BE/DYS/AC**
Sensitivity (%)	100	100
Specificity (%)	100	100

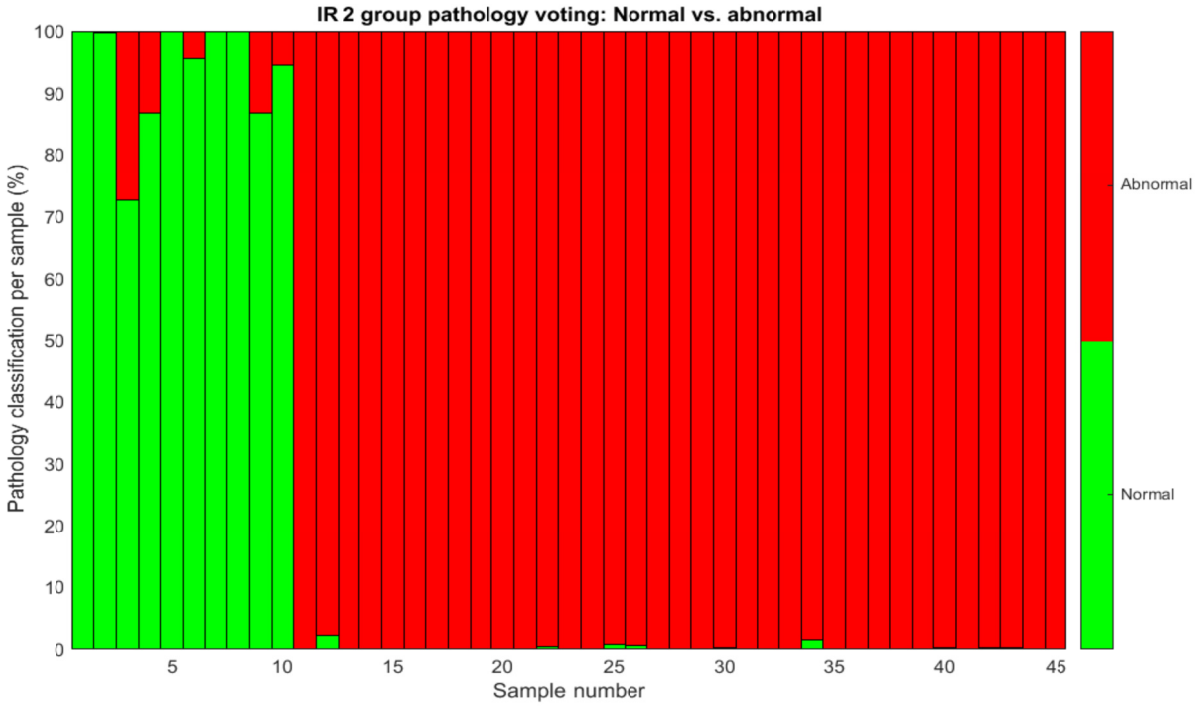


Figure 4-9 Classification of each sample using voting classification for FTIR mapping 2 group model. The proportion of pixels classified as each pathology is shown for each sample in turn. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma. Samples 1-10 are NSQ, samples 11-45 are 'abnormal' i.e. BE/DYS/AC. Every sample is classified correctly.

4.9.2.2. 'Low-risk' versus 'high-risk'

For Barrett's surveillance biopsies, it may be sufficient simply to identify the 'high risk' samples containing dysplasia/adenocarcinoma, in order to identify patients who may require therapeutic intervention, classifying Barrett's and normal squamous together as 'low risk'. The results of this 2 group model for individual spectra are shown in Table 4-8, and Figure 4-10 below.

Table 4-8 Performance of the 2 group model NSQ/BE vs. DYS/AC. Results for individual spectra are shown in A) confusion matrix (with correctly classified spectra in **bold**), and (B) sensitivity and specificity. Threshold voting classification of whole samples is shown in C) confusion matrix and D) sensitivity and specificity. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology	
Predicted spectral pathology	**NSQ/BE (no. of spectra)**	**DYS/AC (no. of spectra)**
NSQ/BE	**295272**	48359
DYS/AC	65176	**328473**

B

	Tissue pathology	
Performance of 2 group model	**NSQ/BE**	**DYS/AC**
Sensitivity (%)	81.9	87.2
Specificity (%)	87.2	81.9

C

	Tissue pathology	
Predicted sample pathology	**NSQ/BE**	**DYS/AC**
NSQ/BE	**17**	1
DYS/AC	5	**22**

D

	Tissue pathology	
Performance of 2 group model	**NSQ/BE**	**DYS/AC**
Sensitivity (%)	77.3	95.7
Specificity (%)	95.7	77.3

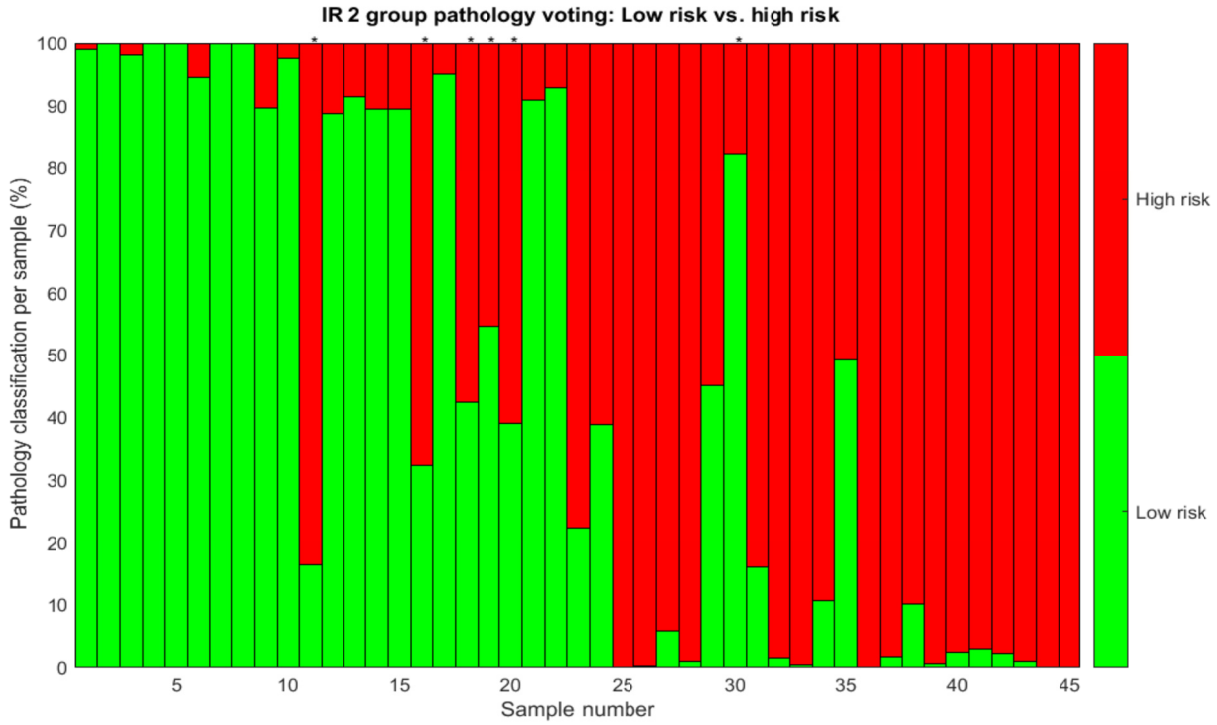


Figure 4-10 Classification of each sample using voting classification for FTIR mapping 2 group model NSQ/BE vs. DYS/AC. The proportion of pixels classified as each pathology is shown for each sample in turn. Samples misclassified using the 30% threshold are marked with an asterisk at the top of the bar. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma. Samples 1-22 are 'low risk' (NSQ/BE), samples 23-45 are 'high risk' (DYS/AC).

The spectral classification performance of the 2 group model is slightly better than the 3 group model in terms of sensitivity for dysplasia/adenocarcinoma, with sensitivity 87.2% (versus 84.6% in the 3 group model). When translated to whole samples using the 30% voting threshold, the sensitivity is identical to the 3 group model at 95.7% (the same sample, D4(2), is misclassified).

A ROC curve can be plotted to identify how the model performance varies with voting threshold (see Figure 4-11). This demonstrates a high performance across a wide range of thresholds, and that using a threshold in the range 15.8-21.0% with this dataset would give a sensitivity of 100%, with a specificity of 77.3%.

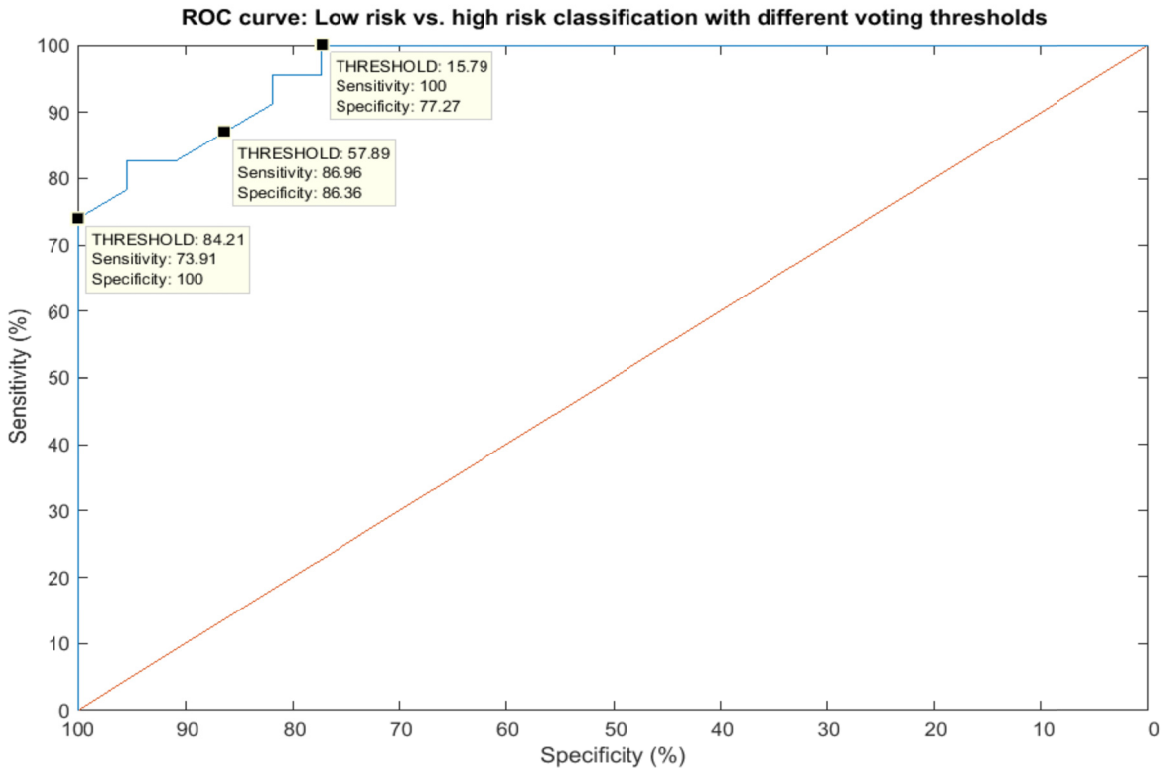


Figure 4-11 Receiver operator characteristic (ROC) curve for identifying 'High-risk' samples. The sensitivity and specificity for classifying whole samples is plotted for voting thresholds from 0-100% (blue line). The red line indicates a reference line of no discrimination.

4.10. Spectral analysis of biochemical differences between pathologies

Spectral analysis was undertaken to identify the underlying biochemical differences between the pathology groups, and the basis for the classification

models described above. Deducing which biomolecules contribute to which peaks can be done using known spectra from the literature or by measuring spectra from pure biomolecules to act as reference standards. At certain wavenumbers there may be overlap with a number of possible bond vibrations (or a number of possible biomolecules) causing spectral peaks at that position. There are also some slight differences in peak wavenumbers quoted in different reference sources. Where they exist, reference standards from human oesophagus are used as the primary source to inform peak assignment for this project. Where human oesophagus reference standards are not available, published data from other tissue sources are used, or 'in house' reference standards. The in house reference standards are a set of pure biomolecules measured previously using FTIR by another member of the Biophotonics Research Unit (Dr M Isabelle). These reference standards are listed in Appendix VIII.

Given the overlap in quoted wavenumbers for certain biomolecules, there may be more than one possible molecular cause for a given spectral appearance, and consequently individual peak assignments listed here are tentative. Nonetheless, where multiple peaks within a spectrum all point to the presence of a particular biomolecule, this is considered strong supportive evidence.

4.10.1. Spectral peak assignment

The first stage of spectral analysis was to examine the key spectral peaks of the mean spectra for each pathology group. The mean spectra are shown in Figure 4-12.

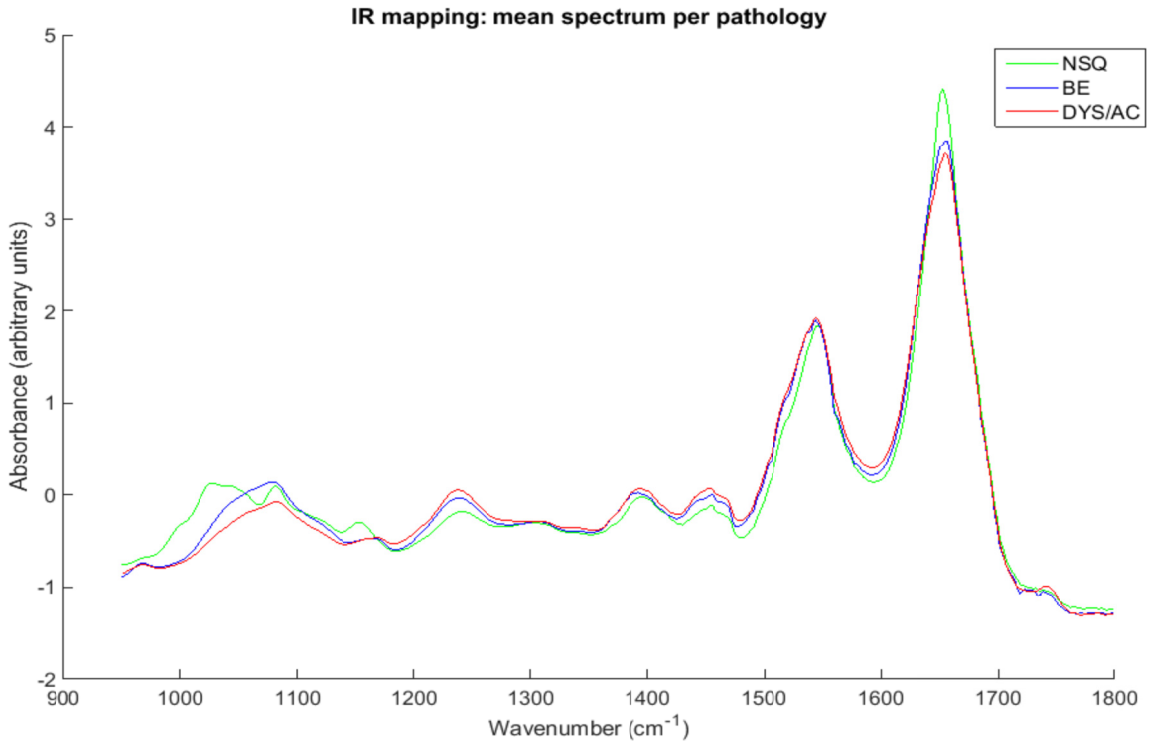


Figure 4-12 FTIR mapping mean spectra. NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/adenocarcinoma.

Whilst a number of subtle differences can be seen between the mean spectra, derivative spectra can be used to visualise the peak positions more clearly, and this is especially helpful when looking at differences between spectra. The second derivative highlights subtle changes to the gradient of the spectrum, and causes effective narrowing of the spectral peaks, allowing greater accuracy in identifying peak position. The second derivative mean spectra for each pathology group are shown in Figure 4-13. The major peaks contributing to each mean spectrum are labelled, and the likely bond vibration and biomolecule causing these spectral peaks are listed in Table 4-9.

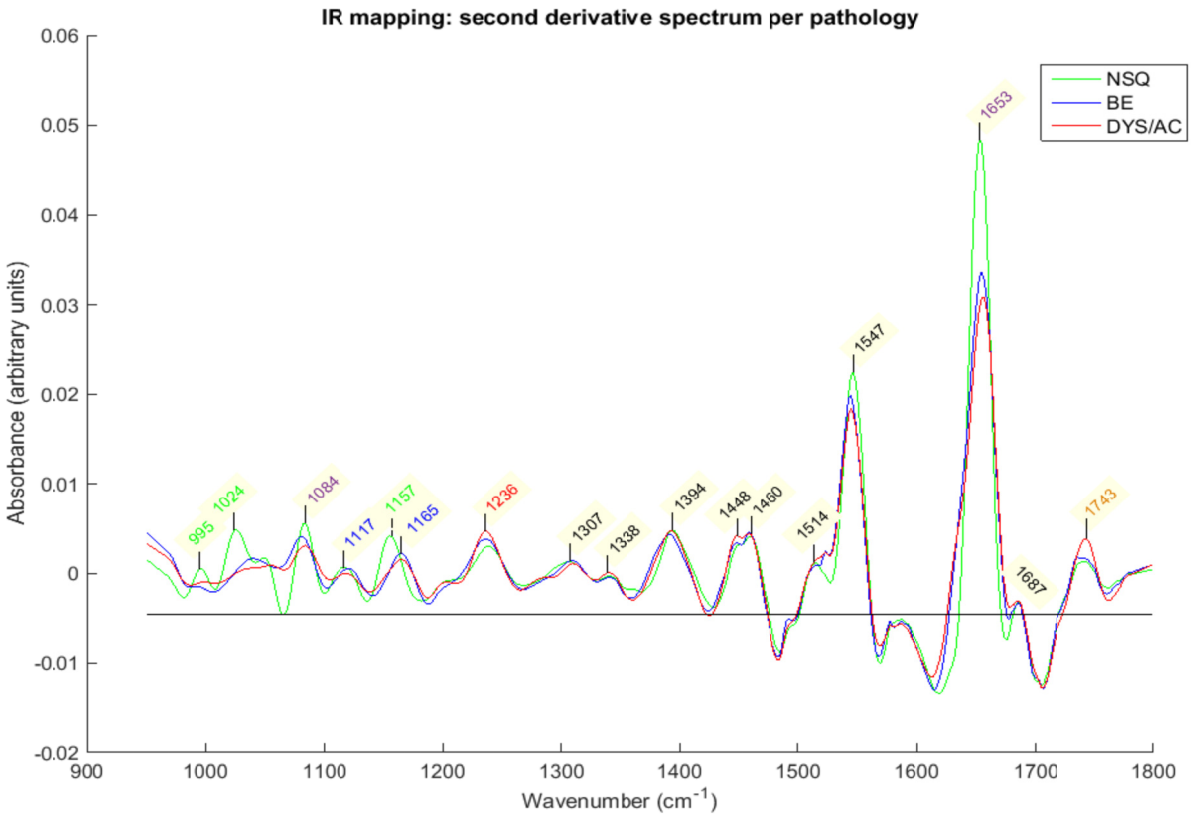


Figure 4-13 Second derivative mean spectra. Peak wavenumber positions are labelled, with the text colour indicating the likely biomolecule mainly responsible for that peak. Green = glycogen, Blue = glycoprotein, red = DNA, Black = amino acid/protein, Orange = phospholipid, Purple = mixed contributions.

By far the largest peak seen in the spectra of all pathology groups is the amide I peak at around 1650cm^{-1}. This is predominantly due to tissue proteins and stretching of the C=O bond in an amide bond, but contributions may also come from glycogen, glycoproteins and DNA (Amrania et al. 2012; Wang et al. 2007). The next largest peak is seen at 1547cm^{-1}, which is the amide II peak resulting from C-N stretch and N-H bend, and is again attributable to tissue proteins. Both the amide I and amide II peaks are higher for the normal squamous group. In previous work in oesophageal tissue by Amrania et al. (2012), the 1650cm^{-1}

peak was considered to be a marker of cytoplasm. The cytoplasm content of normal squamous cells is higher than glandular cells in Barrett's, which in turn is higher than in dysplastic/adenocarcinoma cells: this is consistent with the 1650cm^{-1} peak height seen in Figure 4-13.

Table 4-9 FTIR peak assignments

Wavenumber (cm^{-1})	**Nearest reference peak (cm^{-1})**	**Bond vibration (where known)**	**Biomolecule**	**Reference tissue**	**Reference**
995	992	-	Glycogen	IHRS	IHRS
1024	1026	C-O-H bend	Glycogen	Oesophagus	Wang et al., 2007, Quaroni et al., 2009
1084	1080	PO_2^- symmetric stretch	DNA	Oesophagus	Wang et al., 2003, Wang et al., 2007
	1080	C-O, PO_2^-	Glycoprotein	Oesophagus	Quaroni et al., 2009
	1081	C-C stretch	Glycogen	Oesophagus	Wang et al. 2007, Quaroni et al., 2009
1117	1117	O-H bend	Glycoprotein	Stomach	Wang et al., 2007
1157	1154	C-O stretch	Glycogen	Oesophagus	Wang et al., 2007, Quaroni et al., 2009
1165	1164	C-C stretch	Glycoprotein	Oesophagus	Wang et al., 2007,
	1171		Glycoprotein	Oesophagus	Quaroni et al., 2009
1236	1234	PO_2^- asymmetric stretch	DNA	Oesophagus	Amrania et al., 2012
	1238	C-N stretch, N-H bend	Protein (amide III)	Oesophagus	Wang et al., 2007, Quaroni et al., 2009
1307	1307	-	Amino acid: phenylalanine	IHRS	IHRS

1338	1337	-	Amino acid: phenylalanine	IHRS	IHRS
1370	1375 (approx.)	-	Glycoprotein	Reference standard	Wang et al., 2007
1394	1398	C-H bend	Protein	Oesophagus	Wang et al., 2007
1448-1460	1454	C-H bend	Protein	Oesophagus	Wang et al., 2007
1514	1520	-	Protein: actin	IHRS	IHRS
1547	1545	C-N stretch, N-H bend	Protein (amide II)	Oesophagus	Wang et al., 2007, Wang et al., 2003
1653	1650	C=O stretch	Protein (amide I) Glycogen Glycoprotein DNA	Oesophagus	Amrania et al., 2012, Wang et al., 2007
1743	1744	C=O stretch	Phospholipid	Reference standards Serum	IHRS, Baker et al., 2014 Sheng et al. 2013

IHRS = In house reference standards;

There are a number of glycogen peaks, most clearly at 995cm^{-1}, 1024cm^{-1} and 1157cm^{-1}. In all of these cases the peak is visibly highest in the normal squamous group, suggesting higher glycogen content in this tissue than the Barrett's or dysplasia/adenocarcinoma. This replicates the previous findings in oesophageal tissue using FTIR (Wang et al. 2003, 2007), and the same finding has been made using Raman spectroscopy (Shetty et al. 2006).

The peak at 1084cm^{-1} may be caused by glycogen, but may also be due to glycoproteins: here the normal squamous peak is slightly higher than the Barrett's peak, which is higher than the dysplasia/adenocarcinoma peak. A

possible explanation would be the higher glycogen content causing the higher normal squamous peak, with the relatively higher glycoprotein content giving the Barrett's group a higher peak than the dysplasia/adenocarcinoma group.

Peaks that may be attributable to DNA occur at 1084cm^{-1}, 1236cm^{-1} and 1650 cm^{-1}. However, at each of these wavenumbers there are other biomolecules that may contribute a spectral peak. In the study by Amrania et al. (2012), the 1234cm^{-1} peak was considered to represent DNA and used as a marker of nuclear material. The 1236cm^{-1} peak is visibly highest in the dysplasia/adenocarcinoma group, consistent with an increase in nuclear material.

The peak at 1743cm^{-1} is likely due to phospholipid, and is highest in the dysplasia/adenocarcinoma group.

There are many further small differences between the spectra: these can be more readily visualised by subtracting the mean spectrum of one pathology group from another. However, caution must be exercised in interpretation of these subtraction spectra where the absolute differences in intensity between groups are very small, and there is significant variation in intensity within each group. Figures 4-14 to 4-16 highlight the differences between the spectra by subtracting one group mean from another.

In Figure 4-14, the normal squamous tissue contains strong peaks at 1022cm^{-1} and 1153cm^{-1} consistent with higher glycogen content. The peak at 1650cm^{-1} representing cytoplasm is also much higher in the normal squamous tissue. The peak in the Barrett's tissue at 1377cm^{-1}, along with the shoulder at 1172cm^{-1}, could be attributable to glycoproteins. The other main glycoprotein peak is usually found at 1080cm^{-1}, but this region also contains glycogen and DNA

peaks. This may possibly explain the shape of the peak seen in Figure 4-14 in the region 1068-1092cm^{-1}, with a higher overall content in the Barrett's tissue (due to glycoproteins) but an indentation due to the higher glycogen content in normal squamous tissue. Taken together, the 3 regions indicate a higher glycoprotein content in the Barrett's tissue, which is consistent with the histopathological findings of a higher mucin content in glandular cells.

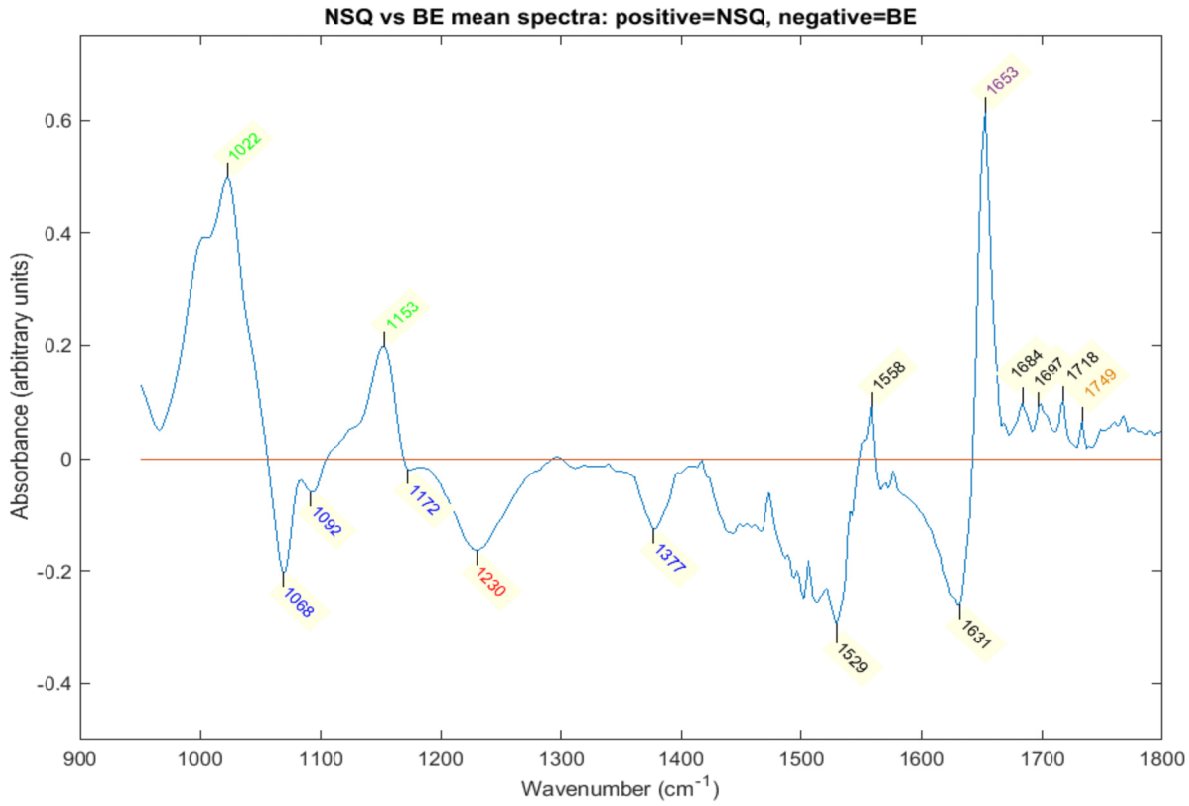


Figure 4-14 Differences between normal squamous (NSQ) and Barrett's (BE) spectra: BE mean spectrum has been subtracted from NSQ mean spectrum, and the peaks (representing higher signal in NSQ tissue) and the troughs (representing higher signal in BE) labelled, with the text colour indicating the likely biomolecule mainly responsible for that peak. Green = glycogen, Blue = glycoprotein, red = DNA, Black = amino acid/protein, Orange = phospholipid, Purple = mixed contributions.

The 1230cm^{-1} peak indicates a higher DNA content in the Barrett's tissue, consistent with an increase in nuclear material. The wavenumber region above 1400cm^{-1} is largely noisy with rapidly oscillating peaks, and consequently is difficult to interpret (with the exception of the large peak at 1653cm^{-1}). The major contributions here are from proteins, and the differences in protein between the groups are hard to disentangle from the spectra.

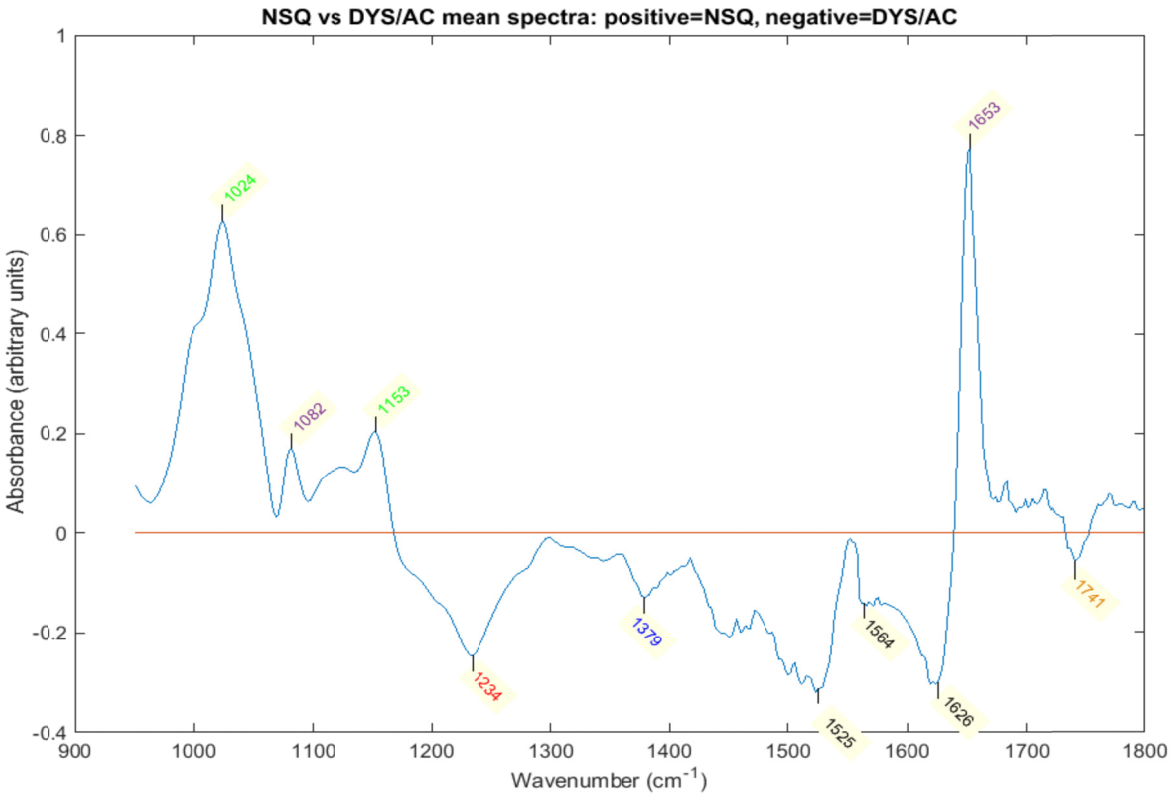


Figure 4-15 Differences between normal squamous (NSQ) and dysplasia/adenocarcinoma (DYS/AC) spectra: DYS/AC mean spectrum has been subtracted from NSQ mean spectrum, and the peaks (representing higher signal in NSQ tissue) and the troughs (representing higher signal in DYS/AC tissue) labelled, with the text colour indicating the likely biomolecule mainly responsible for that peak. Green = glycogen, Blue = glycoprotein, red = DNA, Black = amino acid/protein, Orange = phospholipid, Purple = mixed contributions.

In Figure 4-15, the glycogen peaks are clearly stronger in the normal squamous tissue, at 1024 and 1153cm^{-1}, as is the cytoplasm content represented by the peak at 1653cm^{-1}. The 1082cm^{-1} peak is stronger in the normal squamous tissue, which is likely due to glycogen content, although glycoproteins may also have an effect here.

The DNA peak at 1234cm^{-1} is much stronger in the dysplasia/adenocarcinoma tissue, consistent with increased nuclear material. The glycoprotein peak at 1377cm^{-1} is stronger in the dysplasia/adenocarcinoma group.

Again, above 1400cm^{-1} the subtraction spectrum becomes noisy and harder to interpret, with many small protein peaks.

The phospholipid peak at 1741cm^{-1} is higher in the dysplasia/adenocarcinoma group.

In Figure 4-16, the Barrett's tissue has stronger signal across the wavenumbers 1000-1150cm-1. The major peaks are at 1074 and 1113cm^{-1}, which are most likely attributable to glycoproteins, but this region may also include contributions form glycogen at 1024, 1080 and around 1150cm^{-1}. The other glycoprotein peak at 1373cm^{-1} is stronger in the Barrett's tissue, as is the peak around 1650cm^{-1} representing cytoplasm.

The major DNA peak at 1236cm^{-1} is stronger in the dysplasia/adenocarcinoma group. The peak at 1338cm^{-1} is likely due to the amino acid phenylalanine.

Once again, the wavenumber region above 1400cm^{-1} is difficult to interpret, with many small protein peaks.

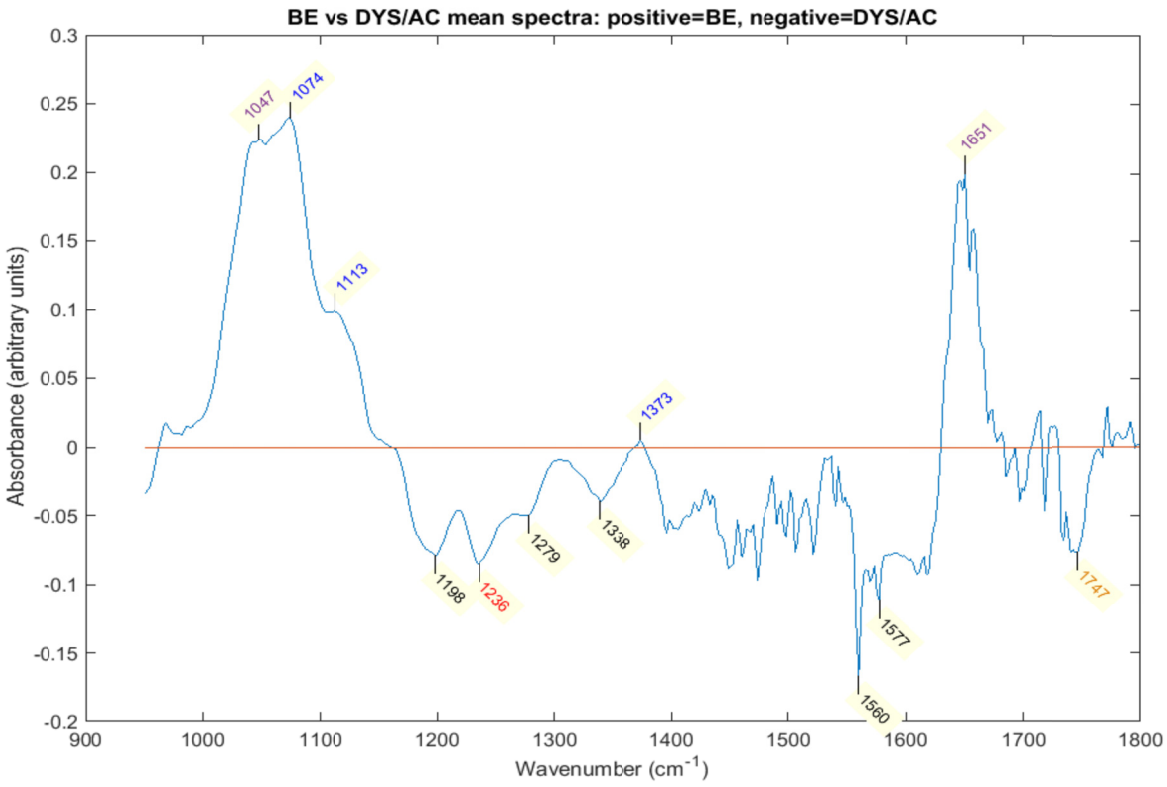


Figure 4-16 Differences between Barrett's and dysplasia/adenocarcinoma spectra: dysplasia/adenocarcinoma mean spectrum has been subtracted from Barrett's mean spectrum, and the peaks (representing higher signal in Barrett's tissue) and the troughs (representing higher signal in dysplastic/adenocarcinoma tissue) labelled, with the text colour indicating the likely biomolecule mainly responsible for that peak. Green = glycogen, Blue = glycoprotein, red = DNA, Black = amino acid/protein, Orange = phospholipid, Purple = mixed contributions.

The peak which stands out at 1747cm^{-1} is again likely to represent phospholipid, with a higher content in the dysplasia/adenocarcinoma group.

Table 4-10 Key discriminatory biomolecules identified in spectra

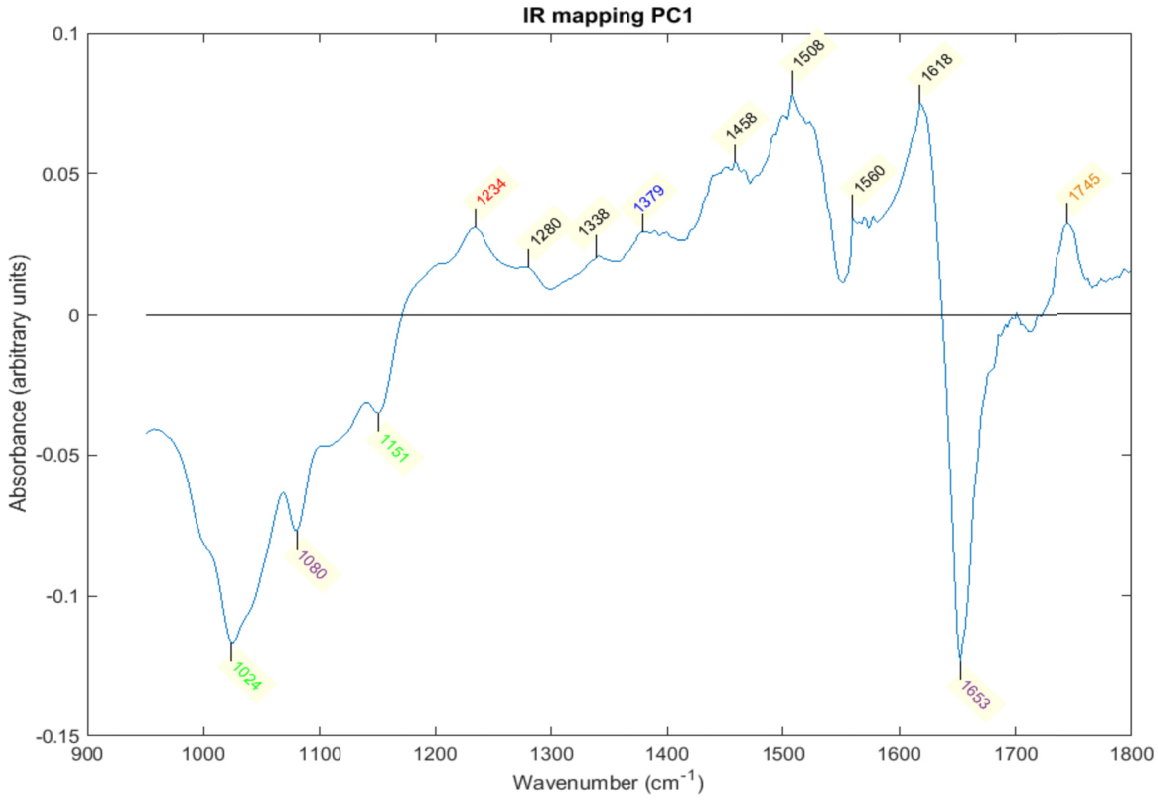
Pathology group	Biomolecule	Cellular component	Wavenumber (cm^{-1})
Normal squamous	Glycogen	Energy store	995, 1022-6, 1153-7
	Protein/glycogen/ glycoprotein	Cytoplasm	1653
Barrett's	Glycoprotein	Mucin	1080, 1113-7, 1171-2, 1373-7
Dysplasia/ adenoca.	DNA	Nuclear material	1234-6
	Phospholipid	Cell membrane/ lipofuscin?	1741-7

4.10.2. Interpretation of principal component loadings

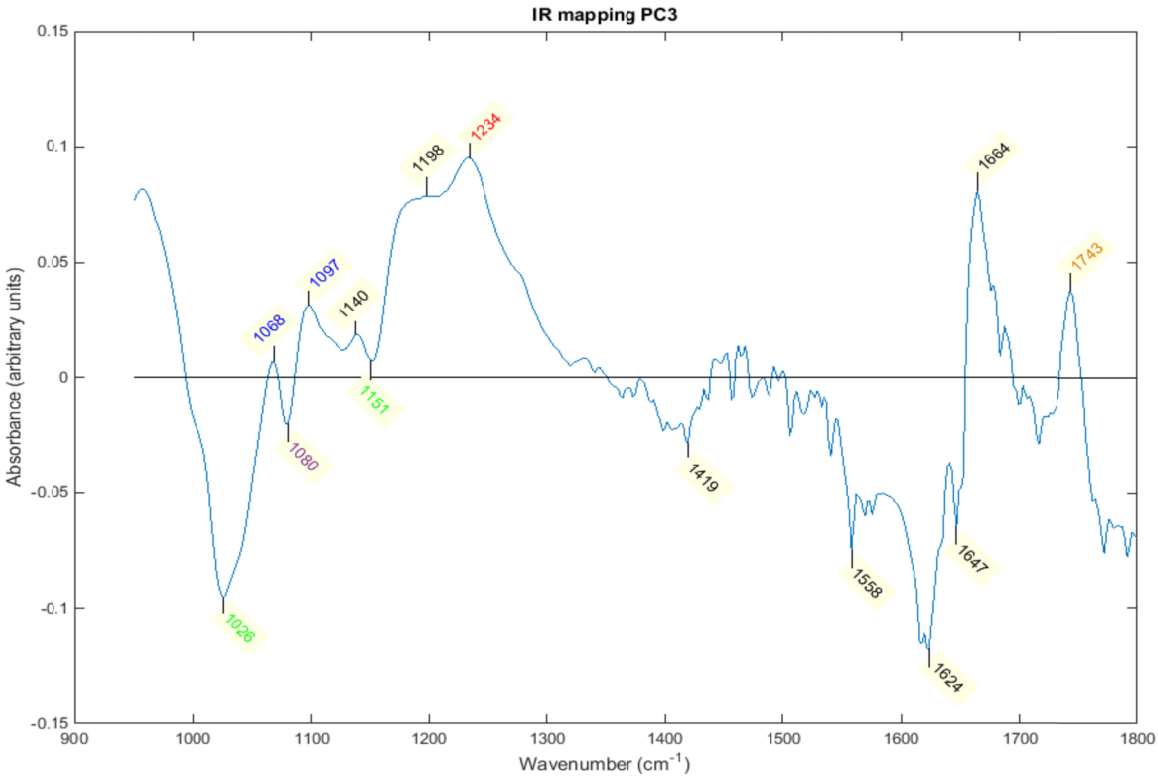
Having analysed the differences between the spectra and the biomolecular contributions to these differences, the principal component (PC) loadings can then be examined to see what spectral differences have been used to discriminate between the groups.

There were 48-50 PCs chosen by ANOVA for each training model: in order to focus on those that contribute the most to the group classification, only the 5 most discriminatory PCs as calculated on F ratio (see Figure 4-4 above) will be discussed here. These PCs are plotted in Figure 4-17 A-E.

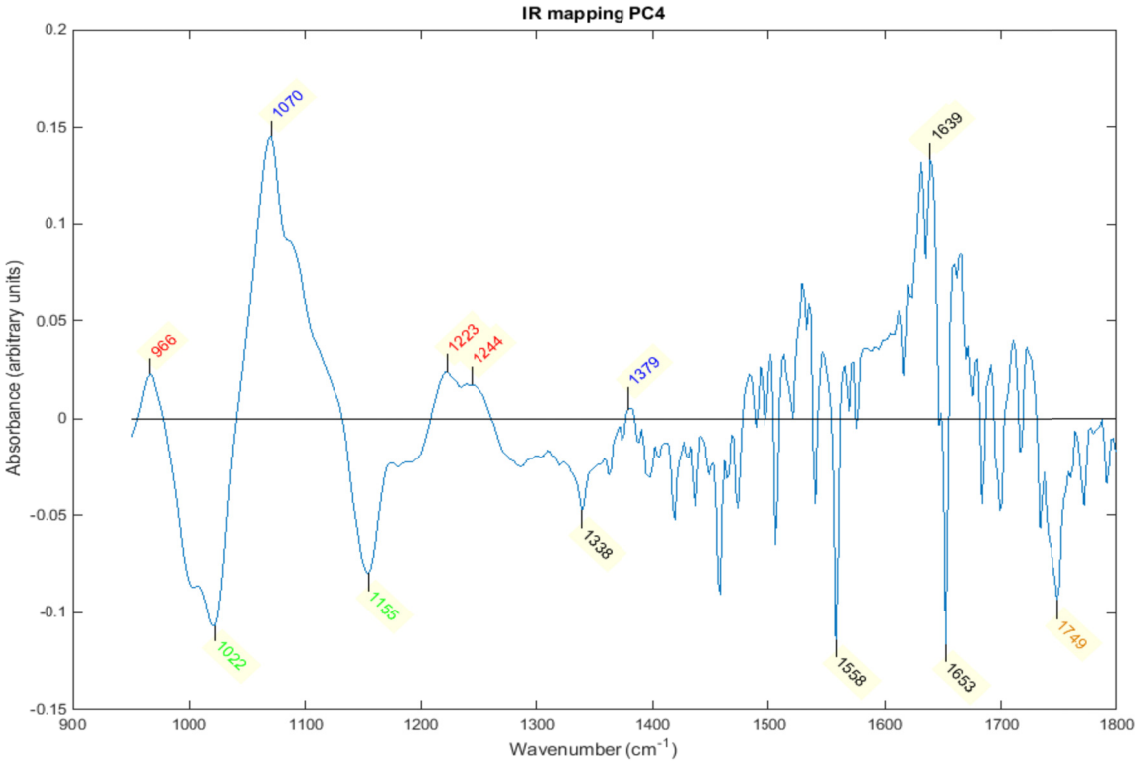
A


IR mapping PC1
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.1
0.05
0
-0.05
-0.1
-0.15
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1024
1080
1151
1234
1280
1338
1379
1458
1508
1560
1618
1653
1745

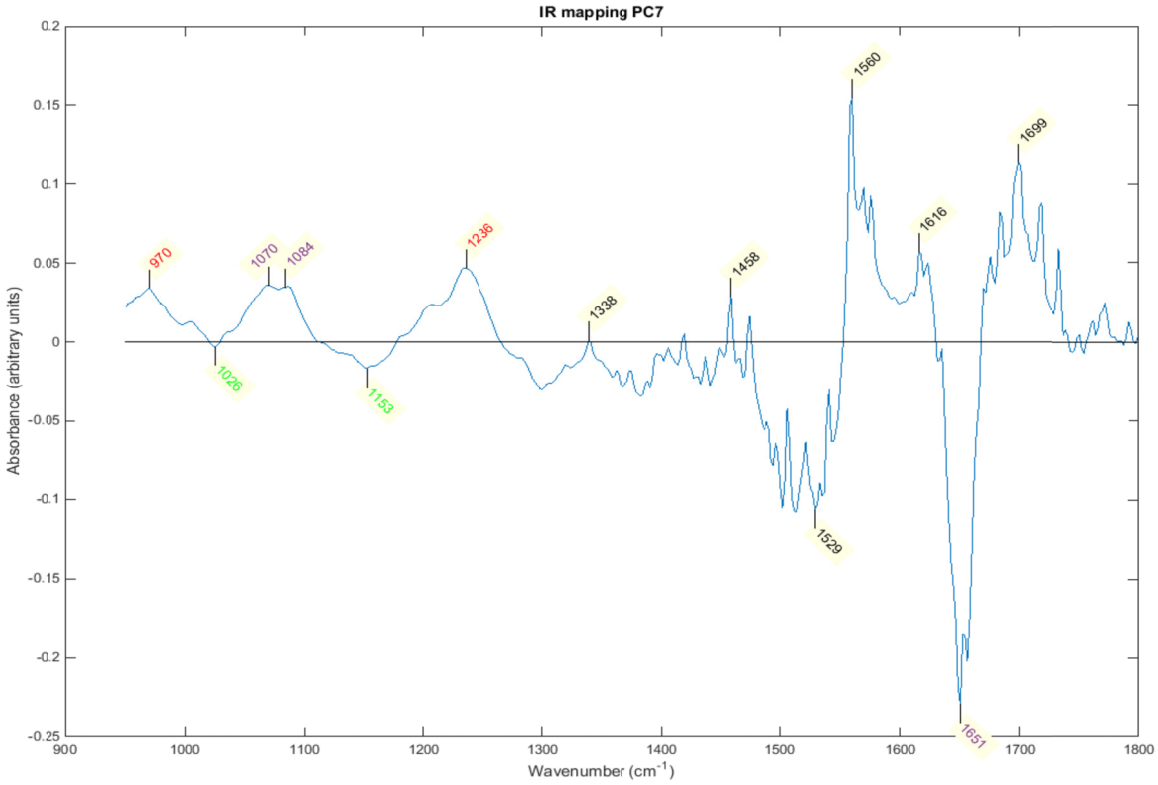
B


IR mapping PC3
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1026
1068
1080
1097
1140
1151
1198
1234
1419
1558
1624
1647
1664
1743

C


IR mapping PC4
Absorbance (arbitrary units)
Wavenumber (cm-1)
966
1070
1022
1155
1223
1244
1338
1379
1558
1639
1653
1749

D


IR mapping PC7
Absorbance (arbitrary units)
Wavenumber (cm-1)
970
1026
1070
1084
1153
1236
1338
1458
1529
1560
1616
1651
1699

E

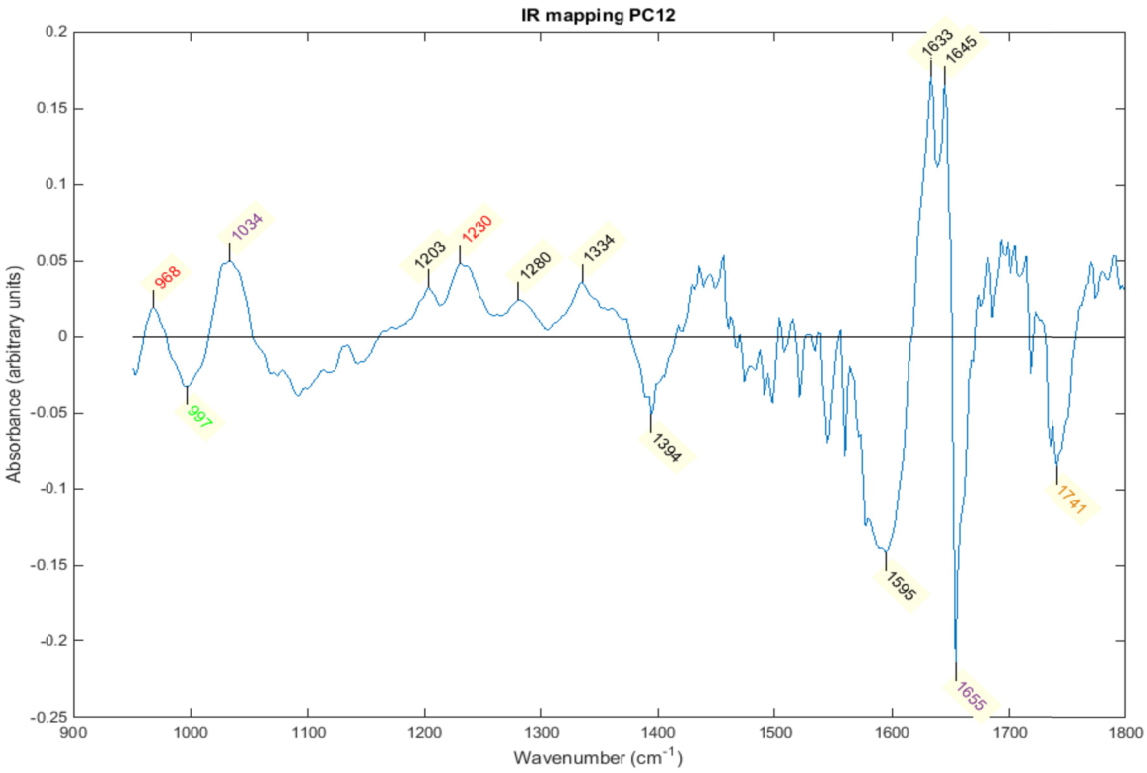


Figure 4-17 Principal component loadings for (A) PC1, (B) PC3, (C) PC4, (D) PC7, (E) PC12. Peak wavenumber positions are labelled, with the text colour indicating the likely biomolecule mainly responsible for that peak. Green = glycogen, Blue = glycoprotein, red = DNA, Black = amino acid/protein, Orange = phospholipid, Purple = mixed contributions.

PC1 is very similar to an inverse plot of the subtraction spectrum for normal squamous versus dysplasia/adenocarcinoma shown in Figure 4-15. It shows (negative) glycogen peaks at 1024, 1080 and 1151cm^{-1}, and a (negative) peak at 1653cm^{-1} representing cytoplasm. There are (positive) peaks at 1234cm^{-1} representing DNA, 1379cm^{-1} consistent with glycoprotein, and 1745cm^{-1} representing phospholipid. It is likely that PC1 separates the dysplasia/adenocarcinoma group (positive) from the normal squamous group (negative).

PC3 shows a strong (negative) glycogen peak at 1026cm^{-1}. The shape of the peaks at 1068-1097cm^{-1} is similar to the normal squamous versus Barrett's subtraction plot: this may be separating glycoproteins (positive) from glycogen (negative). The effect of the glycogen peak at 1151cm^{-1} indents a larger protein peak at 1198cm^{-1} (and creates a smaller peak at 1140cm^{-1}), which merges with the strongly positive DNA peak at 1234cm^{-1}. The positive peak at 1743cm^{-1} represents phospholipid. The region above 1400cm^{-1} is largely noise. Taken as a whole, PC3 appears to separate Barrett's/dysplasia/adenocarcinoma (positive) from normal squamous (negative).

PC4 shows positive peaks at 966 and 1223-1240cm^{-1}. The 1223-1240cm^{-1} peak is likely to have a large contribution from the DNA peak at 1234cm^{-1}. The 966cm^{-1} peak was not clearly seen in the subtraction spectra, but previous work in oesophageal tissue has identified a DNA peak at 970cm^{-1} (Wang et al. 2007), so it is possible this represents a further DNA peak. There is also a strongly positive peak at 1070 and 1086cm^{-1}, which could be due to glycoproteins, though this shows none of the indentation previously seen with an opposing effect from glycogen. There is a small positive peak at 1379cm^{-1} which may also be due to glycoprotein. The strongly negative peaks at 1022 and 1155cm^{-1} represent glycogen. Again, therefore, it seems that PC4 separates Barrett's/dysplasia/adenocarcinoma (positive) from normal squamous (negative).

PC7 contains positive peaks at 970, 1070-1088 and 1236cm^{-1}. The 970cm^{-1} peak may represent DNA as discussed for PC4, as does the 1236cm^{-1} peak. The 1070-1088cm^{-1} peak is harder to interpret, since there is no other supportive evidence that it points to glycoproteins without a positive peak at

1373cm^{-1}. The negative peaks at 1026 and 1153cm^{-1} represent glycogen, and the strongly negative 1651cm^{-1} represents cytoplasm. Therefore PC7 probably separates dysplasia/adenocarcinoma (positive) from Barrett's/normal squamous (negative), although its separation from the Barrett's group on this basis may be weak.

PC12 again shows positive peaks at 968 and 1230cm^{-1} suggesting DNA content. The 997cm^{-1} peak may possibly represent glycogen, though the absence of other glycogen peaks weakens this explanation. The 1034cm^{-1} peak does not appear in the subtraction spectra and it is not clear which groups this peak would separate. There is a further positive peak at 1334cm^{-1} which most likely represents phenylalanine. There is a negative peak at 1394cm^{-1} representing proteins, and the higher wavenumbers are again largely noise. There is a negative peak at 1741cm^{-1} which probably represents phospholipid. Given the DNA peaks, PC12 probably weakly separates dysplasia/adenocarcinoma (positive) from Barrett's/normal squamous (negative), though the negative phospholipid peak goes against this.

4.11. Discussion

This study has demonstrated the ability of FTIR mapping to discriminate between normal squamous, non-dysplastic Barrett's and dysplasia /adenocarcinoma tissue samples. To our knowledge, this is the first study to achieve this using tissue from all four major pathology groups in Barrett's oesophagus. We have used a novel method of analysis to combine spectral classifications across an FTIR map for whole sample predictions. This is also

the first study to analyse fresh-frozen oesophageal sections using high-resolution transmission FTIR mapping.

The key biomolecular differences detected in this study are summarised in Table 4-10 above. Glycogen acts as an energy store and is contained in the cytoplasm. Normal squamous tissue has more abundant cytoplasm than the other tissue types, and the increased metabolic activity of the other tissue types would be expected to reduce the concentration of this energy store. Thus our finding of greater glycogen signal in normal squamous tissue is consistent with tissue biology. This finding has been made previously in studies of oesophageal tissue with both FTIR (Wang et al. 2003, 2007) and Raman spectroscopy (Shetty et al. 2006).

Glandular cells have a higher mucin content than normal squamous cells, with the highly mucinous goblet cells seen in intestinal metaplasia being one of the defining features of Barrett's oesophagus. Therefore our spectroscopic finding of higher glycoprotein signal in Barrett's samples is consistent with this tissue feature, and replicates the previous finding of Wang et al. (Wang et al. 2007).

One of the key histological features used to identify dysplastic or cancerous tissue is that of enlarged nuclei, with a decreased cytoplasm to nuclear ratio. Consequently the finding of stronger DNA signal within this group, suggesting increased nuclear material, is in keeping with the histological basis for discrimination between the groups. This finding has also been reported previously in both FTIR and Raman studies of oesophageal tissue (Shetty et al. 2006; Wang et al. 2007).

Thus the key biomolecular differences found on spectral analysis are consistent with previous findings and correspond to the histological differences between

the tissue types. However the stronger phospholipid signal seen at 1743-7cm^{-1} in the dysplasia/adenocarcinoma group has not been reported previously in studies of oesophageal tissue. FTIR measurements of malignant cells and tissues have reported differing results on the signal strength at this wavenumber. A study of oral cancer tissues reported a markedly lower phospholipid signal at 1745cm^{-1} in the malignant versus normal tissues (Mackanos and Contag 2010). Wang et al. found a lower overall tissue signal in malignant gallbladder tissue at 1740cm^{-1} (Wang et al. 2010), but noted an increased signal at the same wavenumber from malignant tissue plasma membrane and individual malignant gallbladder cells. In a further cell study, signal at 1740cm^{-1} was found to be a marker of apoptosis in a human colon cancer cell line (Gao et al. 2015).

These reported results are somewhat contradictory, but it is conceivable that Barrett's dysplastic or cancerous cells could have increased apoptosis, which in turn could lead to accumulation of the degradation product lipofuscin. Lipofuscin has been posited as a possible marker of dysplasia in work using autofluorescence by DaCosta et al. in both oesophagus and colon (DaCosta et al. 2005; Kara et al. 2004, 2007). This hypothesis is rather speculative but may be worth investigating with future work.

These spectral differences have contributed to the principal components selected as the most discriminatory between pathology groups, and used to develop a classification model. The PC loadings suggested strong discrimination between the normal squamous samples and the other tissue groups, with weaker features to discriminate between Barrett's and dysplasia/adenocarcinoma groups. This correlates with the findings of the

classification models, both for individual spectra and whole samples, which produced a very highly accurate classification of normal squamous samples, and a moderately good classification of Barrett's versus dysplasia /adenocarcinoma groups.

The aims of this study were twofold – firstly to investigate the ability of FTIR mapping to develop a classification model for oesophageal tissue, and secondly to identify key spectral differences to inform work on FTIR measurement and classification of oesophageal cells.

Using the 30% threshold voting system, our 3 group model identified normal squamous samples with 100% sensitivity and specificity. The sensitivity for detecting dysplasia/adenocarcinoma was 95.6%, with only one out of 23 samples being misclassified. Applying the 30% threshold improved the sensitivity of the dysplasia/adenocarcinoma group at the expense of the Barrett's group, which had a sensitivity of 75%.

The 3 group model was tested further using LOPOCV, since between patient differences can contribute significantly to the spectral differences seen in studies with small numbers of patients. Although there was a slight drop in the sensitivity of Barrett's samples, the normal squamous and dysplasia/adenocarcinoma groups were minimally affected. This is further evidence that our 3 group model is highly promising in this small study.

As a more simple test to identify all 'abnormal' samples, the 2 group and 3 group models suggest this could be achieved with very high accuracy indeed, since both models achieved 100% sensitivity and specificity. Alternatively, a 2 group model may be used to identify 'high-risk' samples in patients with known

Barrett's oesophagus: our model was able to achieve 100% sensitivity with specificity 77.3%.

The ultimate goal of developing a classification model for oesophageal tissue would be for use in clinical practice as an automated histopathology tool. Future work, including how these results could be translated into clinical practice are outlined in the summative discussion.

The results in this study could have been affected by a number of inaccuracies in tissue identification and measurement. Histology classification may be inaccurate, as is well-documented, however the use of consensus pathology should limit this problem, and care was taken to select homogeneous regions of pathology for measurement. Consequently, the tissue regions included here are likely to be more homogeneous than most regions chosen at random, and therefore may overestimate the ability of the model to classify sample regions.

Misidentification of regions of the unstained section is possible, and could result in inclusion of tissue in the training model that did not represent the appropriate pathology, or the inclusion of non-epithelial (e.g. stromal) tissue. This inclusion of inappropriate regions in the training model would have impaired its performance.

There was considerable discrepancy in the reporting of the dysplastic samples, with 4 out of 5 patient samples being graded differently as either HGD or LGD by the two pathologists. Since dysplasia is a spectrum of disease, LGD samples may be considered intermediate between non-dysplastic and high-grade dysplasia (HGD). It is possible that any classification model based on tissue biochemistry would find LGD samples more difficult to separate from Barrett's

oesophagus than HGD, and this could result in a significantly poorer classification performance.

Instrument calibration and background measurements were performed prior to every sample measurement, so should be a minimal source of variation. The number of spectra recorded per pixel (64) was selected as a balance between spectral quality and time taken (both for experimental purposes and for any prospect of clinical translation). Using fewer spectral measurements in clinical practice could impair the classification ability. This could be explored further with experimental work, or alternatively modelled using computer simulated degradation of the current dataset.

Further discussion of future work, and possible steps required to translate our findings into clinical practice, are covered in the Summative Discussion (Chapter 10).

CHAPTER 5: FTIR ANALYSIS OF OESOPHAGEAL CELLS

5. FTIR ANALYSIS OF OESOPHAGEAL CELLS

5.1. Preliminary work

Preliminary work on this study began as a collaboration with a group led by Professor Max Diem at Northeastern University in Boston. This group has experience in infrared spectroscopy of cells in a variety of organ systems. We recruited a small number of patients (10 patients, 8 with Barrett's oesophagus, 2 dysplasia) and collected samples which we sent to the Diem group to do initial work in measuring oesophageal cells using an infrared spectrometer. We then performed cytological classification of a number of these cells to develop a training model for pathology classification. These preliminary results have now been published (Townsend et al., 2015).

The project in this thesis was designed as a much larger study with samples from normal squamous, Barrett's oesophagus, dysplasia, adenocarcinoma and stomach cells, and more rigorous inclusion criteria for developing a training dataset.

5.2. Ethical approval and informed consent

Ethical approval was obtained for this study and all patients participating in the study provided informed consent. The ethical approval and consent processes were identical to those mentioned in section 4.1 above, and copies of the Ethical Approval and consent forms are contained in Appendix IV-VI.

5.3. Sample selection

Developing a robust training dataset for spectral classification requires reliable pathology classification of any samples included. Classification of oesophageal cells presents some different challenges from histopathology for tissue samples. Firstly, oesophageal cytology is performed less commonly than tissue biopsy, and expertise in interpretation is correspondingly more limited. Secondly, cells of different morphological appearance may be collected from a region with a homogeneous tissue pathology e.g. normal squamous cells are frequently seen in addition to glandular cells in samples taken from regions of Barrett's metaplasia. It is not known whether these cells would have spectral appearances consistent with normal squamous cells taken from healthy mucosa, or signal more similar to glandular Barrett's cells.

In order to be certain that cells included in the training dataset were representative of the correct pathology, a biopsy was taken from the area after cell collection, and only those cells whose cytological appearance was consistent (on review by 2 cytopathologists) with the contemporaneous endoscopy and biopsy results were included. In cases where the cytology and histology results did not agree, the cells were not included in the training model.

For later analysis of whole samples included in the test dataset, the classification of the whole sample was based on the endoscopy and biopsy result as the gold standard. To remove further the possibility of misclassification, patients were only included in the test dataset if they had no history of more advanced disease i.e. to be included as a normal control they must have no history of Barrett's oesophagus, and to be included as Barrett's

oesophagus they must have no history of oesophageal dysplasia/adenocarcinoma.

5.4. Sample collection

Samples were collected from patients undergoing scheduled endoscopy, either as part of Barrett's surveillance or indicated for other clinical grounds. In order to enrich the sample population for patients with oesophageal adenocarcinoma, some samples were also collected from patients undergoing surgery for oesophageal cancer. Normal oesophageal squamous cells used as controls were collected from patients undergoing endoscopy for symptoms of dysphagia or dyspepsia in whom no endoscopic or histological abnormality was identified. Samples of stomach cells were taken from the fundus of the stomach in order to test the ability to differentiate stomach cells from oesophageal glandular cells.

Cytology samples were collected at endoscopy using an endoscopic cytology brush (Medi-Globe, Achenmule, Germany) passed down the instrument channel of an endoscope, under direct vision at endoscopy. The region of interest was brushed for around 3-5 seconds. The cytology brush was then withdrawn into its sheath, the brush removed from the endoscope, and the brush itself immediately cut off the long wire introducer and placed into a specimen pot containing formalin (10% neutral buffered formalin, Genta Medical, York, UK) (later in the project this was changed to ThinPrep CytoLyt^{TM} solution (Hologic Inc., Massachusetts, USA), the reasons for this are discussed in section 5.5.2.1 below). The brush containing the cells was then stored at room temperature until slide preparation.

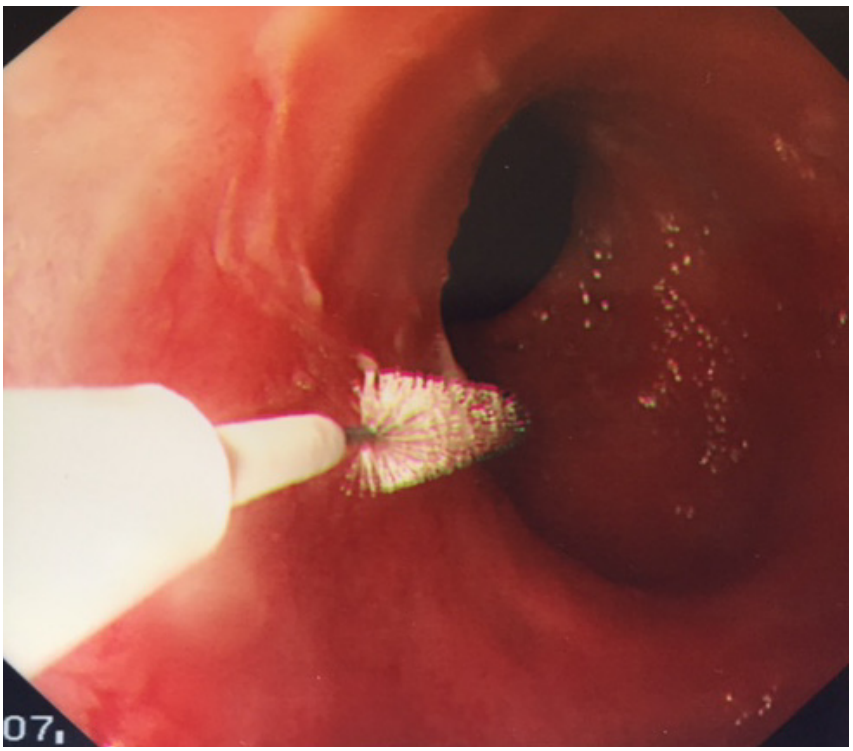


Figure 5-1 Collection of cell sample at endoscopy

For the samples collected at surgery for oesophageal cancer, the same technique was used but without passing the brush through an endoscope. The brush was immediately placed in a formalin-filled specimen pot as above.

After cell collection at endoscopy or surgery, tissue biopsies were taken from the region of interest and sent for histology, to allow comparison of the histology and cytology findings.

5.5. Cell sample preparation

5.5.1. Initial protocol

Since no one in the group currently performs experimental work with cells I set up a link with the cytology department at Cheltenham General Hospital to

support this project, and enable me to prepare the cell slides with initial supervision from a cytology technician.

The initial protocol used for cell preparation was based on a protocol used by the group in Boston with whom we collaborated in the preliminary work.

To ensure the collected cells were effectively transferred from the cytology brush to the formalin solution, the specimen pot was vortexed for approximately 30 seconds prior to removal of the brush. The brush was then disposed of, and the cell suspension transferred to a Sterilin pot (Sterilin Ltd., Newport UK) and topped up to approximately 1ml with formalin. The Sterilin pot was then centrifuged at 3000 revolutions per minute (rpm) for 5 minutes. The supernatant was then discarded, and the cell pellet diluted to 1ml with 0.9% NaCl. This solution was then vortexed, and then centrifuged at 3000rpm for a further 5 minutes. The supernatant was then discarded and the pellet diluted a second time to 1ml with 0.9% NaCl. This suspension was vortexed and then centrifuged at 3000rpm for a third time for 5 minutes. The supernatant was again discarded, and the pellet diluted to 1ml NaCl. This final suspension was vortexed, and then 6 drops (approximately 0.3ml) were taken by pipette and placed in the Cytospin conical funnel (EZ Single Cytofunnel with white filter cards, Thermo Scientific) (see Figure 5-2). The Cytofunnels were loaded with a calcium fluoride slide (Crystran 1mm x 20mm Raman grade polished window) onto which the cell sample would be deposited. Calcium fluoride slides were used as they have very few spectral peaks in the biological region of interest. The Cytospin (Thermo Scientific) was then run at 1500rpm for 3 minutes: the centrifugal force acts on the cell suspension in the conical funnel, which is driven through a 5mm hole onto the slide. Excess liquid is absorbed by the filter paper, leaving a

deposit of cells on the calcium fluoride slides. The slides were left to air dry for 10 minutes, and then stored in a sealed container at room temperature until measurement.

Figure 5-2 Cytofunnel (Thermo Scientific). The calcium fluoride slide is loaded into the slide holder, which is then closed with the filter paper over the slide. The cell suspension is placed into the conical funnel, which is closed with a lid, then placed into the Cytospin.

5.5.2. Optimising cell preparation protocol

A first batch of 10 cytology samples were used to test the above protocol, and the slides reviewed with a cytopathologist. Of these, 5 samples were considered

to be too thick, with densely clumped cells, and 5 were considered to be good samples, with a relatively even distribution of cells.

Following this, the final dilution stage was changed to dilute the cell pellet to 2ml with normal saline rather than 1ml, and place 3-6 drops into the cytospin conical funnel depending on whether the sample was judged to be highly cellular or appeared blood-stained on visual inspection. A second test batch of 12 samples was then prepared. Of these, one slide broke in the cytospin centrifuge, one sample had very few cells, one was too heavily bloodstained, and the other 9 were considered good samples. On discussion with the cytopathologist, it was felt that some variability was inevitable, and we decided to continue using this protocol for the remainder of the samples.

Later in the project, a number of further modifications were made to the protocol to address problems that became apparent with the slides. Some slides had large pieces of non-cellular debris, and an additional step was introduced, after the slides were fully dry, to wash them in distilled water.

It later became apparent that many samples had small numbers of cells, so the final dilution stage was adjusted again, such that the final dilution of the cell pellet was now 0.5-2.0ml NaCl, depending on sample thickness.

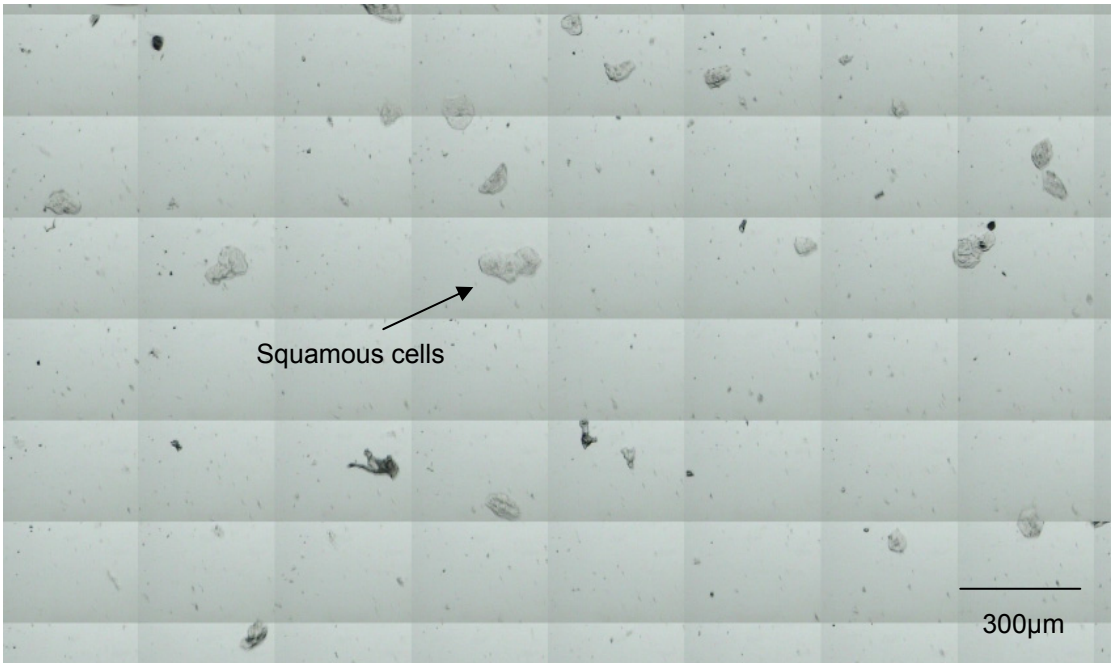
High numbers of red blood cells continued to be a problem, and on advice from a cytopathologist we tried storing samples in ThinPrep CytoLyt™ solution (Hologic Inc., Massachusetts, USA) rather than formalin. CytoLyt™ is a methanol based buffered preservative, designed to have a haemolytic effect. An experiment to compare the spectral effects of this change is presented in section 5.5.2.1.

5.5.2.1. Comparison of CytoLyt™ and formalin fixatives

We designed an experiment to compare CytoLyt™ and formalin as fixatives for storing and preserving the cells. Samples were collected from 5 patients undergoing endoscopy: for each patient 2 brushings were taken from the same region of the oesophagus, with one sample stored in CytoLyt™ and the other in formalin. The samples were then prepared using an identical protocol, and the resulting slides visually compared. In general, the samples stored in CytoLyt™ had slightly fewer red blood cells, and there was slightly less clumping of cells. An example of this is shown in Figure 5-3.

However there was considerable variation in the quality of the slide preparation, which was not systematically attributable to either CytoLyt™ or formalin. The most striking example of this was in sample 61A and 61B, shown in Figure 5-4. These samples were collected from the same region of the oesophagus in the same patient, at the same time, and prepared in the same way (the only difference being the storage solution). The cell preservation and cell density in 61A is far greater than in 61B, where there is a small deposition of cells at the circumference of the cell spot only. Since this finding was not a consistent difference between the two fixatives, it seemed rather to illustrate the variability of slide quality using the current technique. Several slides prepared later in the project (with either CytoLyt™ or formalin) showed a similar appearance.

A



B

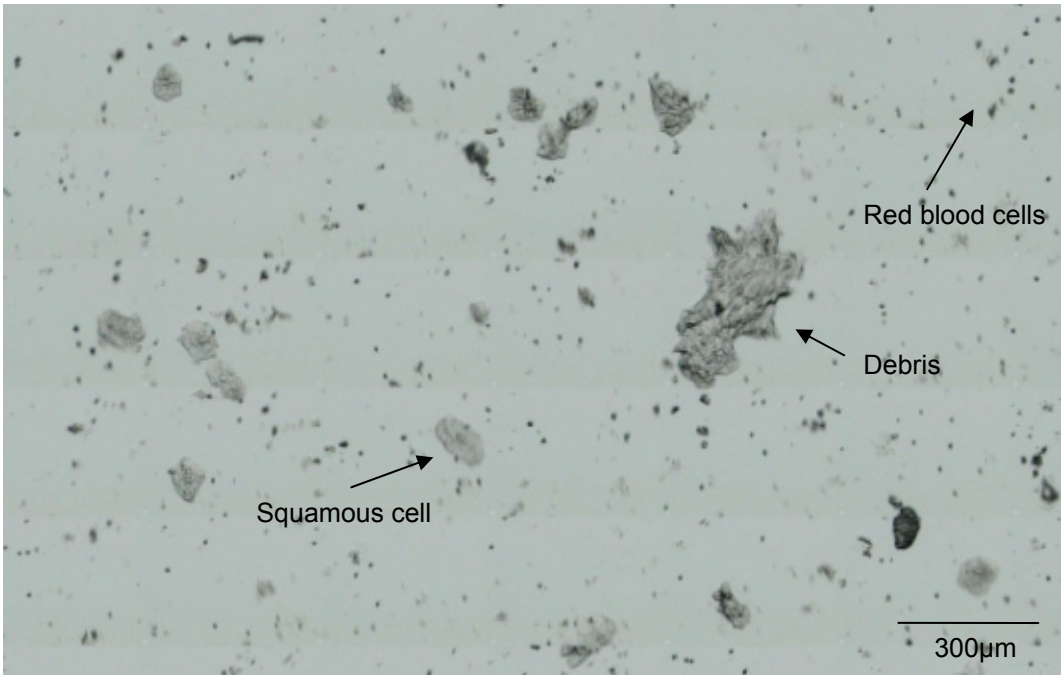


Figure 5-3 Comparison of cell slides A) stored in CytoLyt™ and B) stored in formalin. Cells were collected from the same region of oesophagus in the same patient at the same time, and slides prepared using an identical protocol.

A

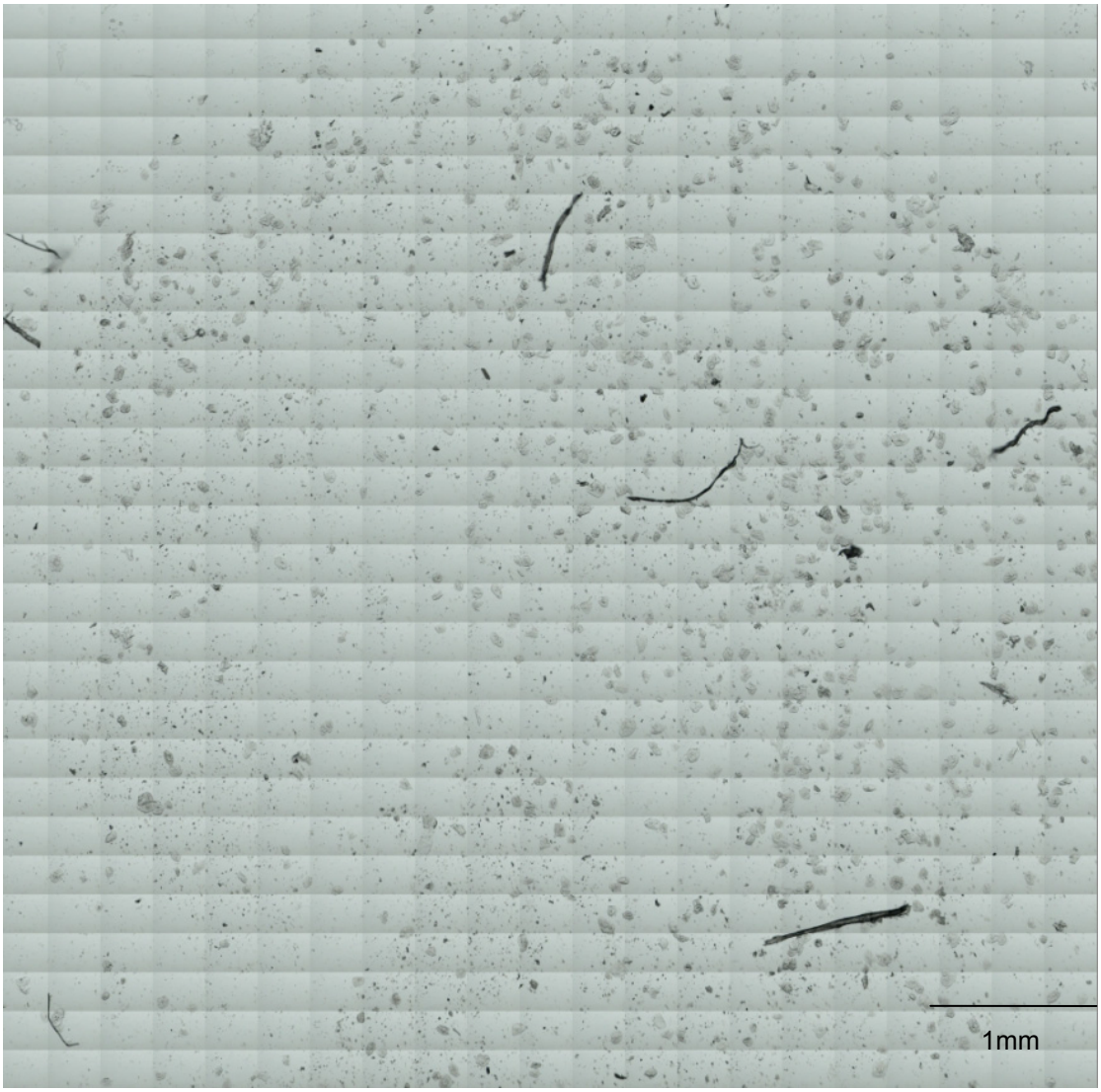


Figure 5-4 A Comparison of cell slides A) 61A and B) 61B. Cells were collected from the same region of oesophagus in the same patient at the same time, and slides prepared using an identical protocol.

B

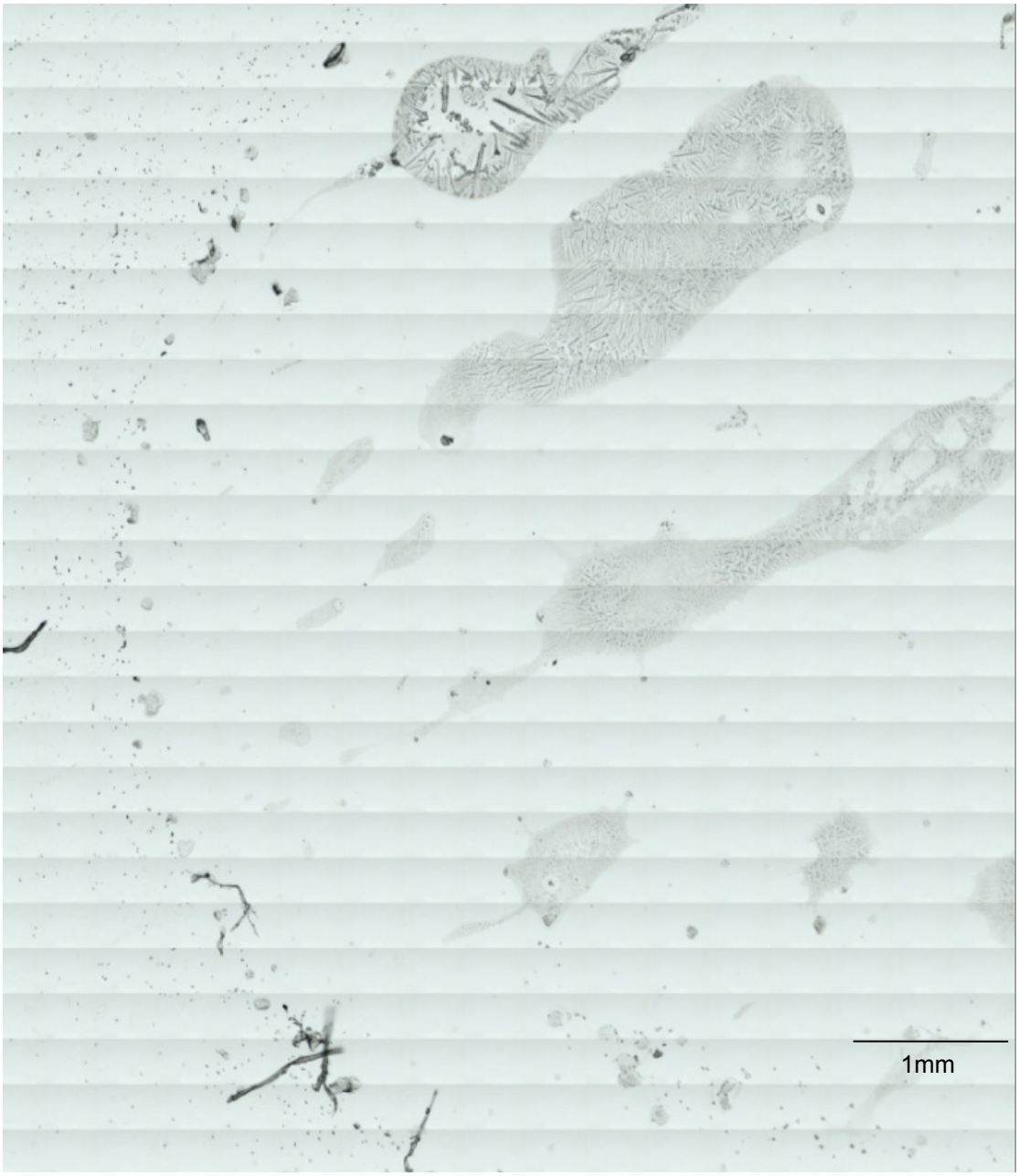


Figure 5-4 B Comparison of cell slides A) 61A and B) 61B. Cells were collected from the same region of oesophagus in the same patient at the same time, and slides prepared using an identical protocol.

Since the CytoLyt™ showed a slightly improved appearance over the formalin fixative, we considered switching to use this fixative for cell preservation.

However, it was important to investigate whether this would have any effect on the resulting cell spectra. We therefore used PCA to see if any spectral features could be identified to discriminate between the samples stored in CytoLyt™ and those stored in formalin. 25 PCs were selected and the most discriminatory identified using ANOVA. The 2 most discriminatory PCs are plotted in Figure 5-5.

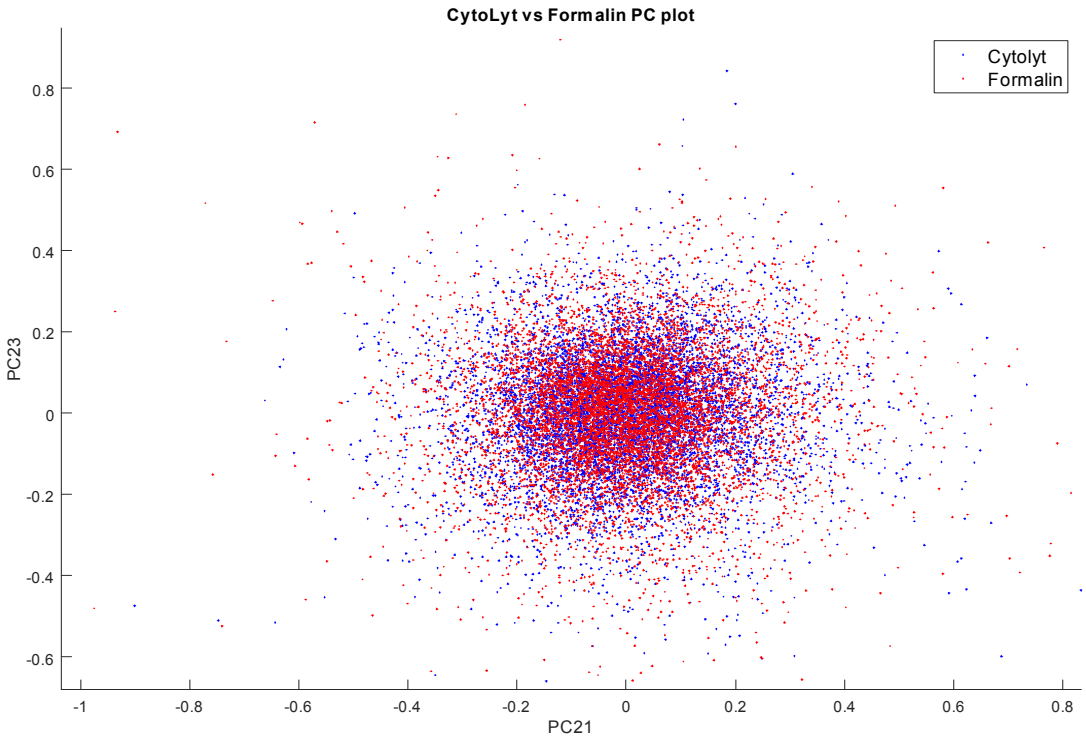


Figure 5-5 PCA plot showing 2 most discriminatory components for comparison of cells stored in CytoLyt™ or formalin.

The most discriminatory principal components were PCs 21 and 23, suggestive of poor discrimination, and the scatter plot in Figure 5-5 shows a seemingly random appearance, with no separation between the two groups. On this basis, we concluded that there was a negligible difference in the contribution to the spectrum from the two fixative solutions. We therefore decided to switch to

CytoLyt™ since it gave a slightly better slide quality, with fewer red blood cells and reduced clumping.

5.6. Sample measurement

5.6.1. Choosing the infrared spectrometer

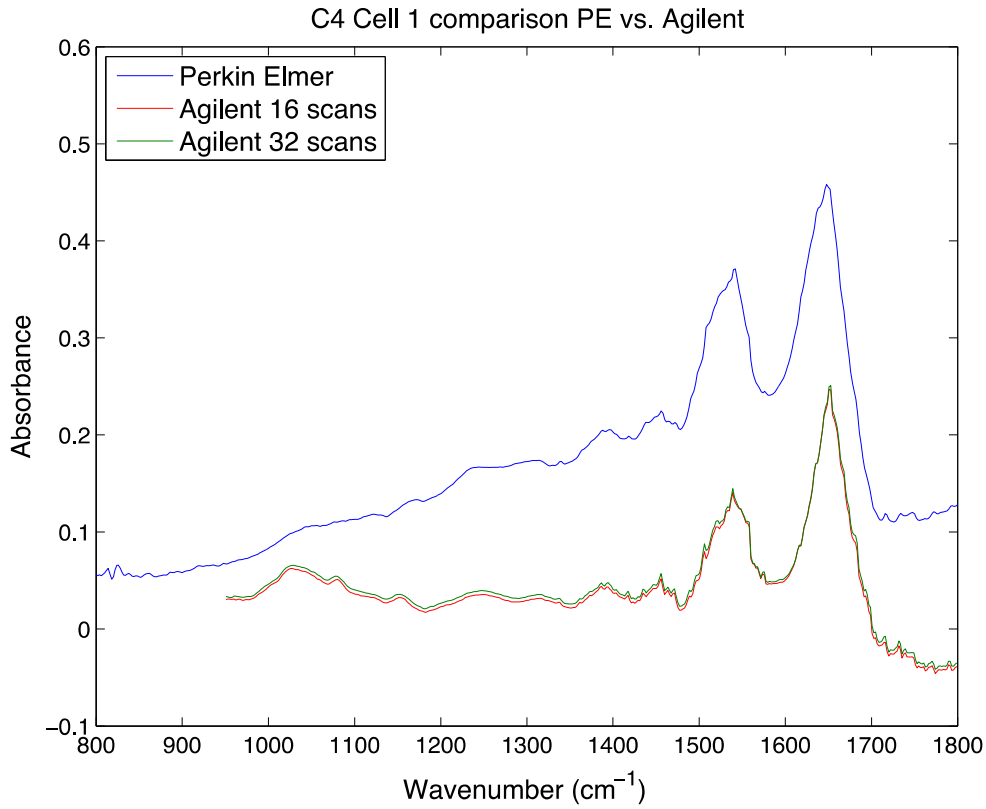
The FTIR tissue mapping study was performed using the Agilent FTIR microspectrometer in Exeter. However, previous FTIR work in cells had been done on a Perkin Elmer Spectrum One FTIR spectrometer at the Biophotonics unit in Gloucester, and it was suggested that this system gave a higher quality spectral signal than the Agilent system.

On both systems multiple scans can be measured per pixel, then the mean spectrum is stored as the measurement for that pixel. Using more scans per pixel improves the signal to noise ratio. Each scan on the Agilent system is of lower quality than the Perkin Elmer, but this can be compensated for by its ability to collect more scans in a similar time frame. In order to decide which system to use for our study, the signals from the two systems were compared.

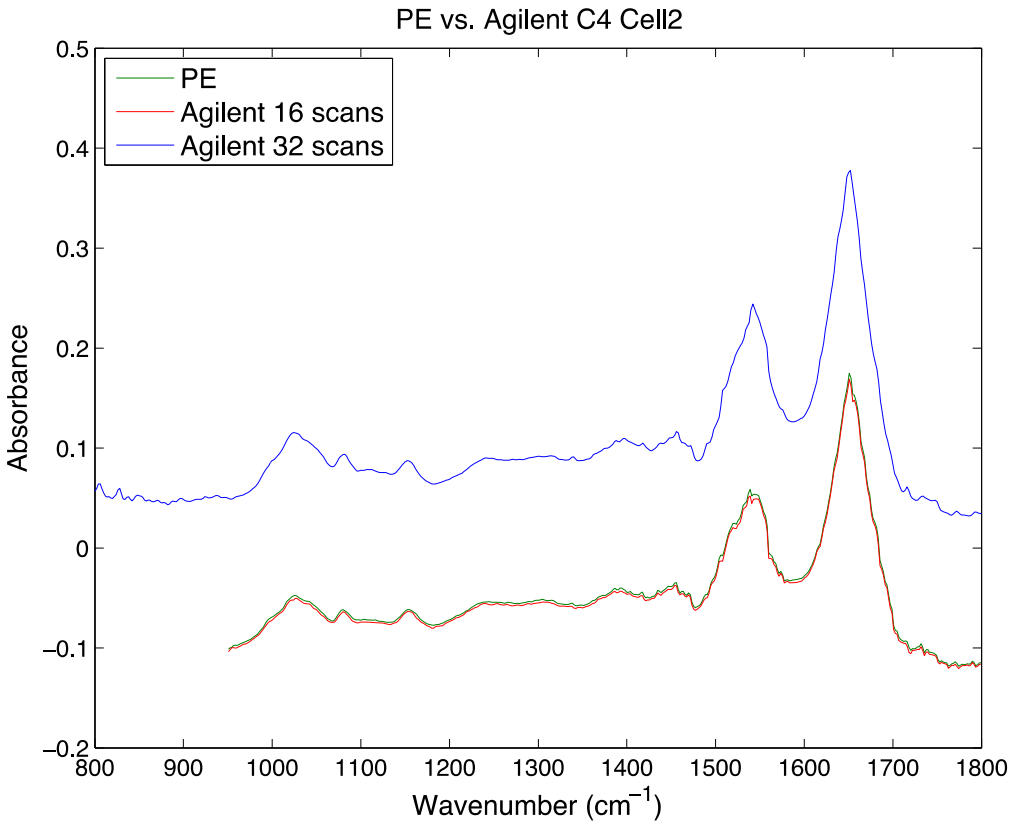
One of the slides of oesophageal cells was chosen, and a small region of the slide selected for mapping on both systems. Comparable parameters were chosen that would be used for mapping samples: PE measurements were performed using 2 scans per pixel, 4cm^{-1} resolution and 6.25μm pixels. For the Agilent, 2 measurements were performed for each cell, at 16 and 32 scans per pixel, both at 4cm^{-1} resolution and 5.5μm pixel size.

The measured region contained 4 cells: the mean spectrum from each cellular region was calculated and plotted in Figure 5-6.

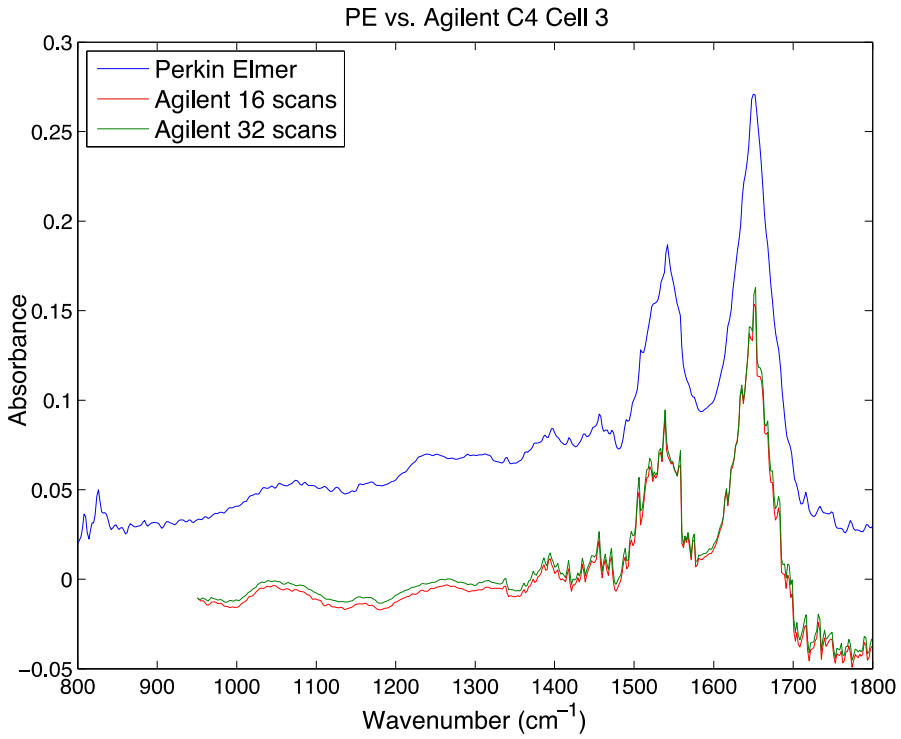
A


C4 Cell 1 comparison PE vs. Agilent
Perkin Elmer
Agilent 16 scans
Agilent 32 scans
Absorbance
Wavenumber (cm^{-1})
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

B


PE vs. Agilent C4 Cell2
PE
Agilent 16 scans
Agilent 32 scans
Absorbance
Wavenumber (cm^{-1})
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

C



D

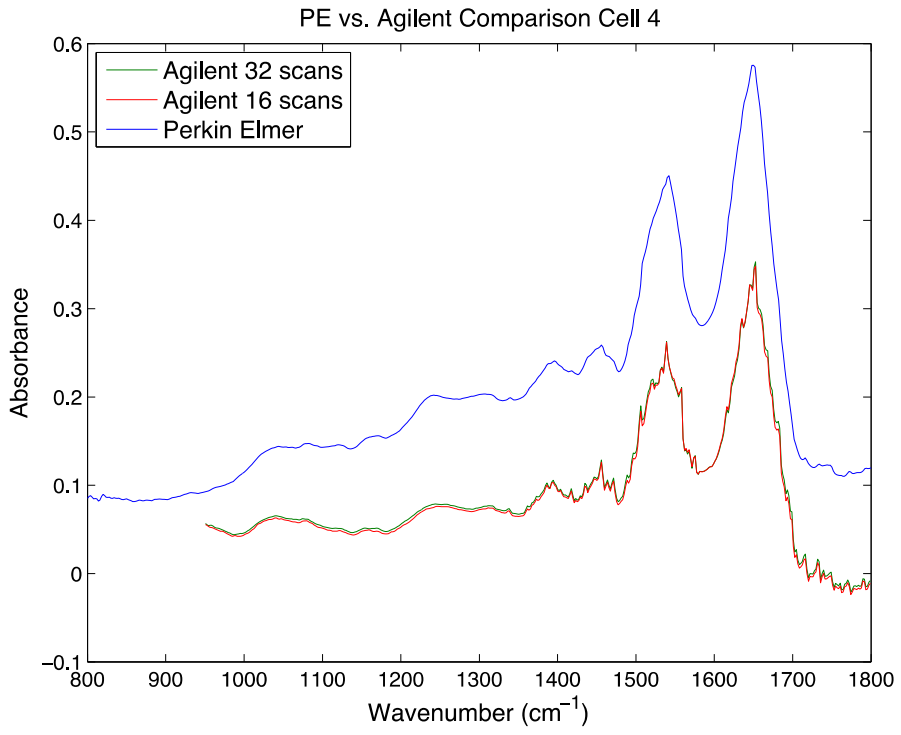


Figure 5-6 Comparison of mean spectra for cells 1-4 measured on the Perkin Elmer (PE) and Agilent infrared spectrometers.

For each of the 4 cells, while the overall spectrum is similar for all 3 measurements, there is more noise in the spectra measured on the Agilent. There is very little difference between the spectra measured using 16 versus 32 scans on the Agilent system. We therefore decided to use the PE system for our measurements, since this gave a less noisy signal using only 2 scans per pixel.

5.6.2. Final protocol used for sample measurement

Samples were measured at the Biophotonics unit at Gloucestershire Royal Hospital using a Perkin Elmer Spectrum One FTIR spectrometer with a Perkin Elmer Spotlight 400 imaging system, containing a 16-element focal plane array detector. Infrared absorption maps were obtained in imaging mode, with measurements in transmission mode, raster scanning at 6.25μm per pixel, at 4cm^{-1} spectral resolution across a wavenumber range of 750-4000cm^{-1}. The automated atmospheric correction for CO_2 and water vapour was turned on for our initial experiments, but this was later switched off as a result of our work to optimise sample measurement and data processing (see section 5.7.3 below). An initial background reading was taken using 120 scans per pixel from an acellular region of the slide. A white light image of the entire cell spot area (5mm x 5mm) was captured. This enabled selection of a region of the cell sample to be measured. We initially planned to measure a 4mm x 4mm region for the Barrett's, dysplasia and adenocarcinoma samples (anticipating that these samples would have fewer cells of the representative pathology than the normal squamous samples), however this was later reduced to a 4mm x 2mm region to reduce the time taken to measure each sample. Since the total time

taken to measure the 4mm x 4mm area (around 8 hours) exceeds the maximum measurement time on the Perkin Elmer system without refilling the liquid nitrogen dewar to maintain adequate detector cooling, the region was subdivided into quadrants or halves for measuring. Maps of infrared absorption were collected across the region using the above settings with 2 scans per pixel.

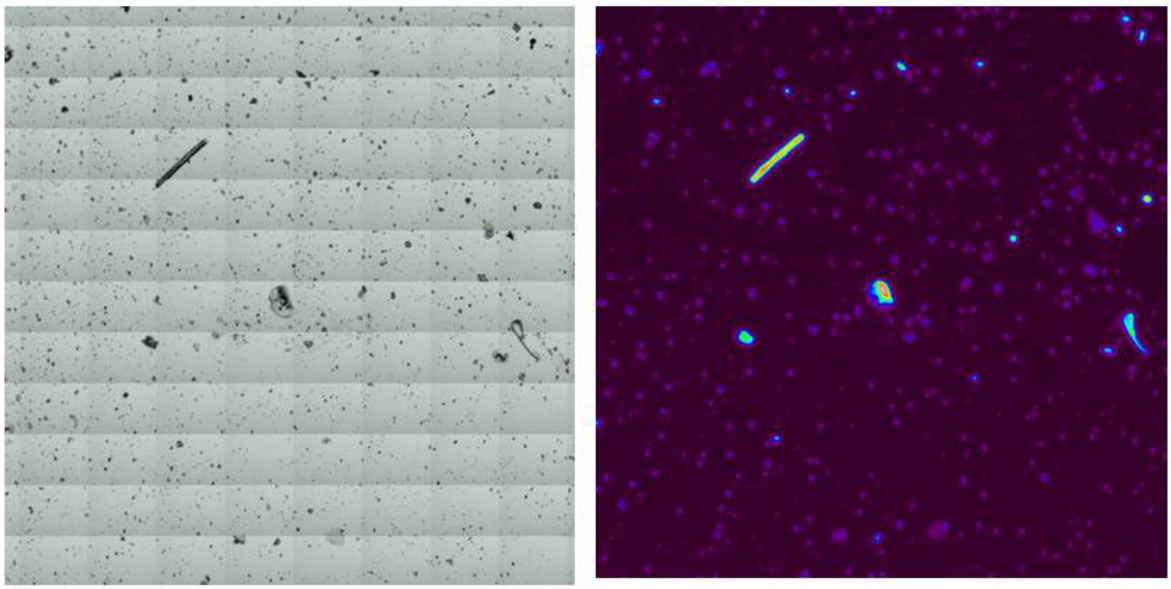


Figure 5-7 White light image of cells and raw absorbance map created

5.6.2.1. Effect of time delay between plating and measurement

Both as part of this study, and if this technique were to be applied in clinical practice, it may be necessary to store cell slides prior to FTIR measurement. We therefore performed an experiment to investigate the stability of the cells and the corresponding spectral measurements over time.

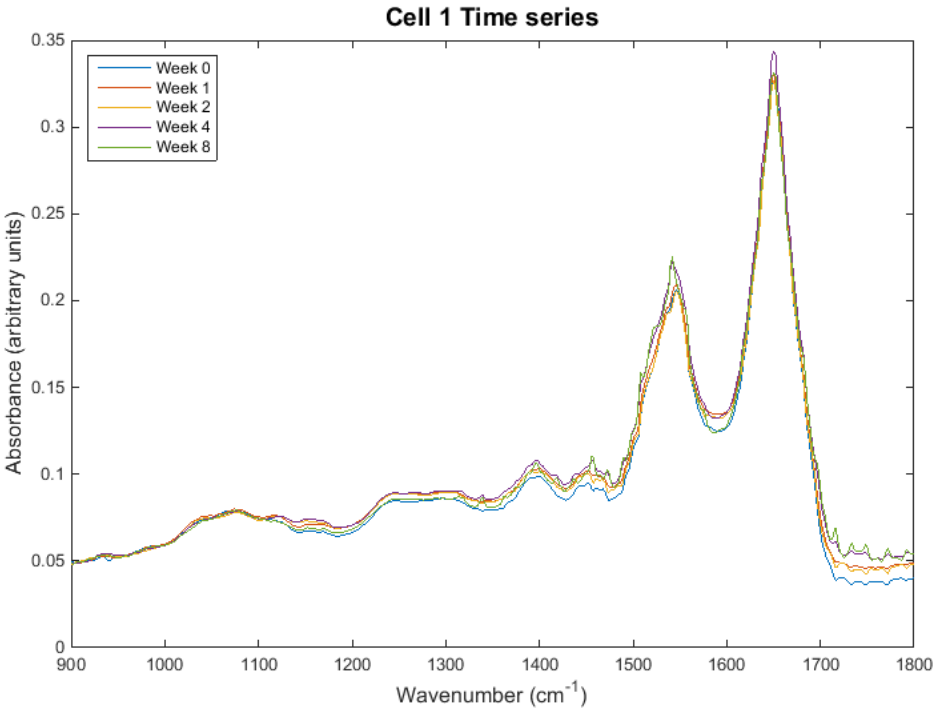
A single cell sample (collected, stored and plated as above) was measured (using the same protocol each time) at a number of time points: immediately after plating (<12 hours), then again after 1 week, 2 weeks, 4 weeks and 8 weeks.

Cells on the slide, and some regions of cell debris identified by the binary mask, were selected and the spectra compared across the various time points. Some examples of the mean spectrum for certain regions at different time points are shown in Figure 5-8.

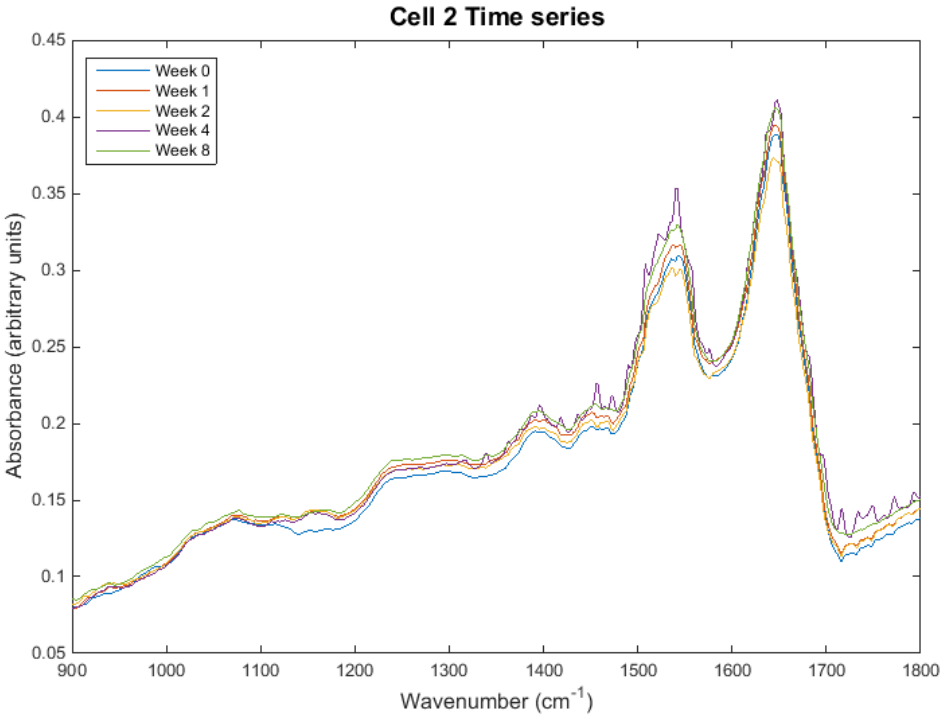
On visual assessment of the spectra, overall there is very little difference between the spectra for a given region measured at different time points, for either cells or cell debris. In Cell 1, Cell debris 1 and Cell debris 2, there is slightly increased noise in the spectra above around 1450cm^{-1} in the Week 8 measurement. This wavenumber region corresponds to the protein region of the spectrum and could thus possibly represent protein changes occurring between weeks 4 and 8. However a similar appearance is seen in Cell 2 at week 4, but is then not seen in the later week 8 measurement. It is possible that this noise is simply within the limits of normal variation for a single cell measurement.

The cells therefore appear stable and give consistent spectral measurements up to 4 weeks after plating. Beyond this, it is possible that some protein breakdown occurs, increasing the noise in the protein region of the spectrum. Consequently cells should be measured within 4 weeks of plating.

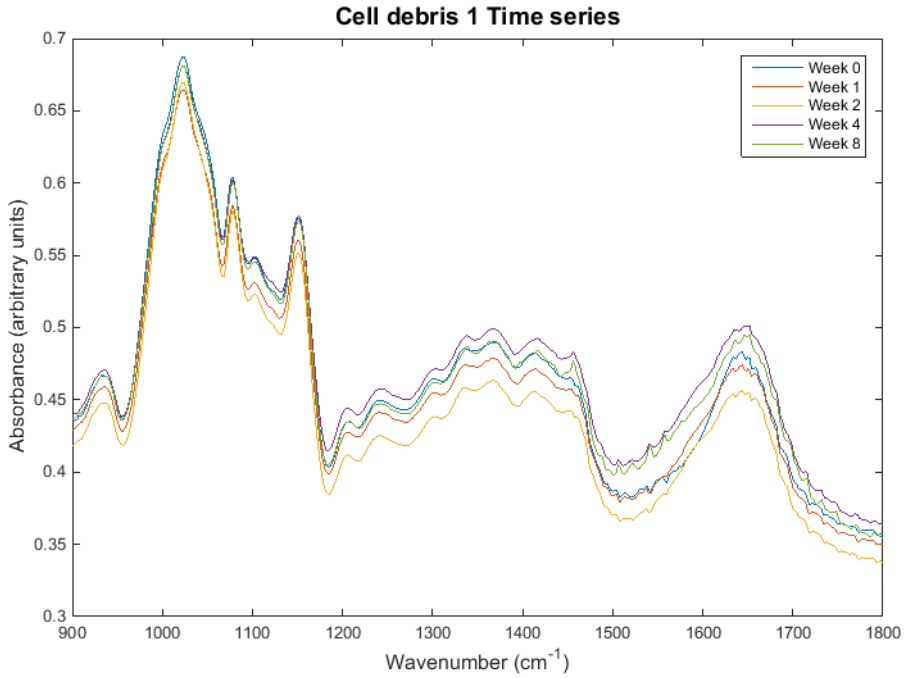
A


Cell 1 Time series
Week 0
Week 1
Week 2
Week 4
Week 8
Absorbance (arbitrary units)
Wavenumber (cm-1)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

B


Cell 2 Time series
Week 0
Week 1
Week 2
Week 4
Week 8
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

C



D

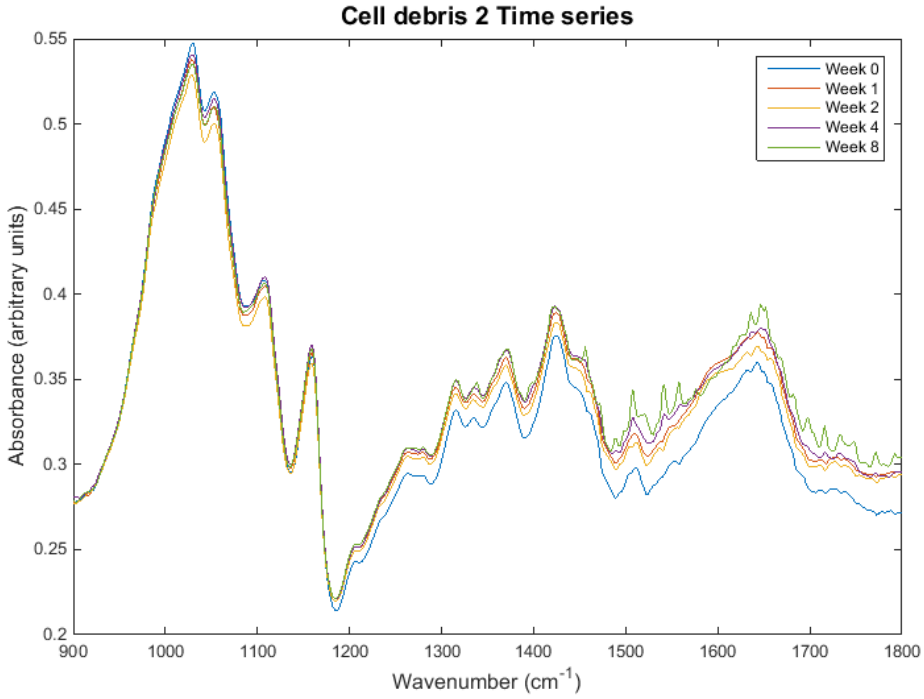


Figure 5-8 Mean spectrum of each cell region measured at different time points. A) Cell 1, B) Cell 2, C) Cell debris 1, D) Cell debris 2.

5.7. Data processing: optimising cell spectra and automated cell identification

The overall aim of data processing was to develop training and test datasets as described in general in section 2.4, and specifically for FTIR mapping in section 4.7 above. The initial data pre-processing steps are similar to those reported for FTIR mapping (section 4.6). The automated cell identification process described in section 5.7.2 below was developed with Dr Gavin Lloyd from the Biophotonics Unit in Gloucestershire Royal Hospital.

Data measured on the Perkin Elmer system are initially recorded as transmittance of infrared and must be converted to tissue absorbance for analysis.

With the relatively smaller signal coming from cells compared to tissue, noise effects due to atmospheric water can be significant, and must be corrected for. This is described in more detail below.

A further step to increase the signal to noise ratio is then applied: a Savitzky-Golay filter. This acts to smooth the data by fitting the curve between a small number of datapoints (we used 11 datapoints) to a low order polynomial (we used 2^{nd} order).

An extended multiplicative scattering correction (EMSC) is then applied, which has several functions. EMSC normalises the data relative to a reference standard spectrum (in our case we used a mean spectrum weighted equally across the pathology groups, calculated by taking the mean value of the mean spectrum for each pathology group), applies a scattering correction (outlined further below), and can be used to identify outlying spectra. Since EMSC

correction uses least squares fitting to a reference spectrum, it gives a measure of distance for each spectrum in the dataset from the reference spectrum. A large distance implies great difference from the reference standard, and this can be used as a threshold to exclude anomalous measurements. We used a threshold that excluded less than 5% of spectra from the dataset.

5.7.1. Mie scattering correction

Mie scattering is a recognised problem in infrared spectroscopy of small particles such as cells. To correct for this, we developed a correction algorithm based on previously published techniques (Bassan et al. 2009, 2010): the technique is described in detail in these papers. Firstly, parameters must be chosen for the range of size of the particle (in this case diameter of the cell, which we estimated to be 10-50μm depending on cell type), and the refractive index of the particle (the authors of the paper describing the technique estimated a value of 1.3 for biological cells, and used a range of 1.1-1.5, so on this basis we used the same values). These parameters are used to plot curves that represent the range of possible scattering effect for particles of this size: in our model we used 1000 curves. Principal components are then used to describe the difference between these curves and the reference spectrum (in this case our weighted mean spectrum). In the original paper 7 PCs described 99.9% of the variance, so we used PCs 1-7 in our study. These PC loadings are then used in a least squares fitting algorithm to estimate the scattering contributions for each spectrum.

5.7.2. Automated identification of cells

Our intention was to develop a system that used mapping to measure large areas and then automatically identify cells within the measured region. In this way an automated measurement system could be used for a cell slide without having to visually identify and measure each cell individually.

We used a post-processing algorithm to detect cellular regions and extract the spectra from this region, assigning it to a particular cell.

The algorithm combines spectral information and size criteria to determine which regions may be cellular. From the FTIR mapping work, we knew the largest protein peak occurred at around 1650cm^{-1}, and that this corresponded well to regions of cytoplasm in our study and previous work. Since all cells should contain cytoplasm, this seemed a good choice of wavenumber to use to identify cells. The algorithm then uses a threshold intensity of the 1650cm^{-1} peak to decide whether a spectrum has come from a cellular region or not: if the 1650cm^{-1} signal is strong enough, the spectrum is considered to be from a cell. Cellular regions are then coloured black, and non-cellular regions white, to create a 'binary mask' representing the cells on the slide. To exclude proteinaceous debris that may also have a strong signal at 1650cm^{-1}, size criteria are used to remove contiguous areas detected by the binary mask that exceed 400 pixels in total area (125μm x 125μm – 125μm is approximately the size of 3-4 large squamous cells). An example of a cell slide and its binary mask are shown in Figure 5-9.

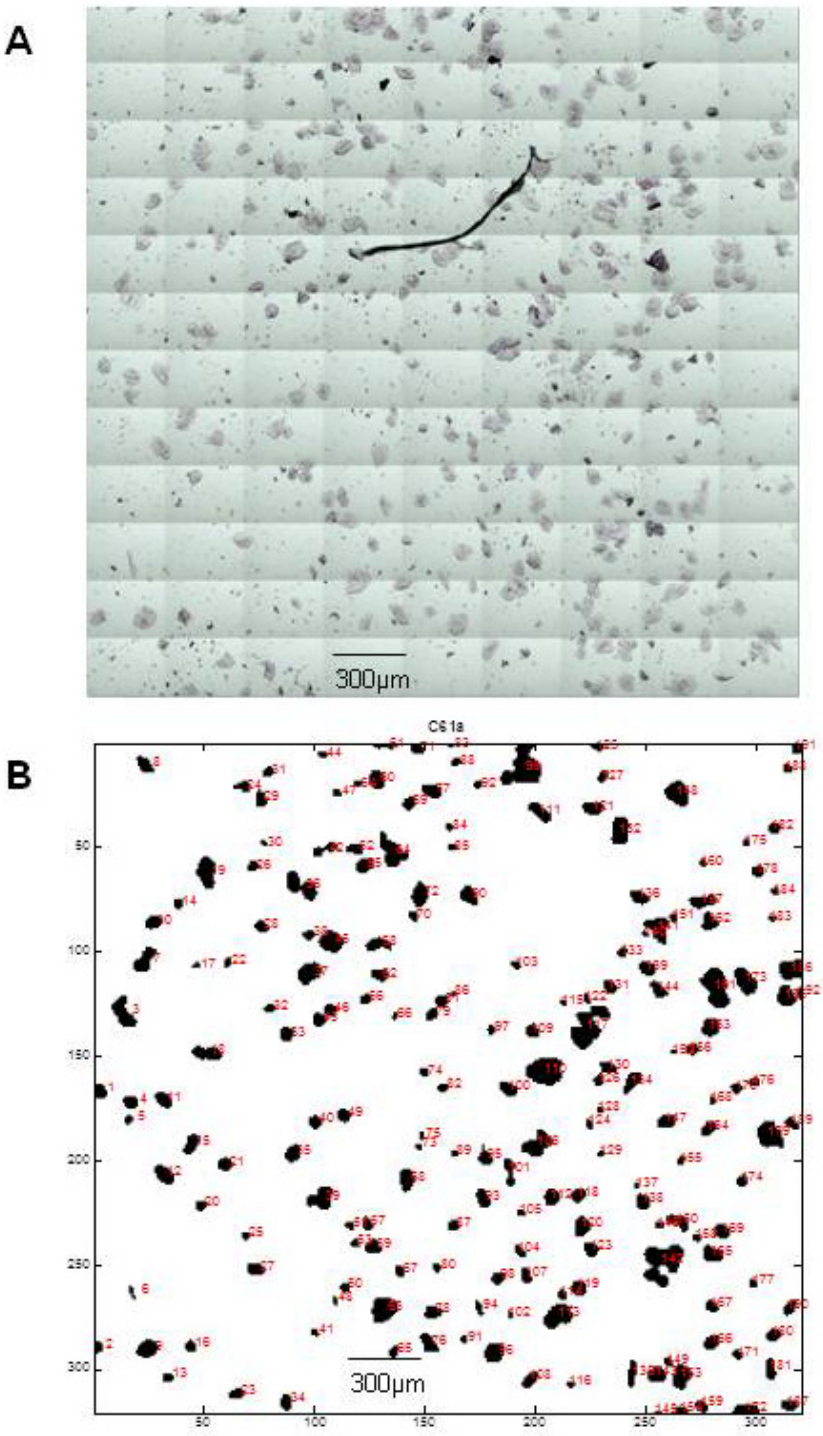


Figure 5-9 A) White light image of an unstained slide prior to measurement B) Binary mask created from infrared map of the same slide, with black regions corresponding to cellular areas detected by the automated algorithm. The binary mask has excluded the large piece of debris on size criteria.

5.7.2.1. Testing the binary mask against point measurement

The 'binary mask' algorithm was used to highlight areas of the map that contained cells, and extract spectra from these regions. It is critical that this process extracts the spectral information corresponding to the signal measured from the correct cell, otherwise our training model would not be composed of the correct spectra.

There are a number of potential sources of error, e.g. if the extracted spectra do not correspond exactly with the boundaries of the cell, incorrect labelling such that the spectral information from one cell is labelled as another cell/region on the slide, or if the binary mask does not accurately identify cellular regions.

To confirm the binary mask was able to perform these functions accurately, a test sample was measured firstly in mapping mode, using the binary algorithm to detect cells, then manually taking point measurements of the same cells, and comparing the spectra.

The same settings were used for both point measurement and mapping: 2 scans per pixel, resolution 4cm^{-1}. A background of 120 scans was collected prior to measurement for both mapping and point mode. The pixel size for mapping is 6.25μm x 6.25μm, and any pixels detected as part of the cell by the binary mask were included; for point measurement an aperture size of 24μm x 30μm was used, which fitted within the cells without including any edge regions or non-cellular areas. For larger cells 2 regions were measured. Figure 5-10 shows the white light image of the unstained cell slide, and the binary mask created from the image measured using FTIR mapping.

A B

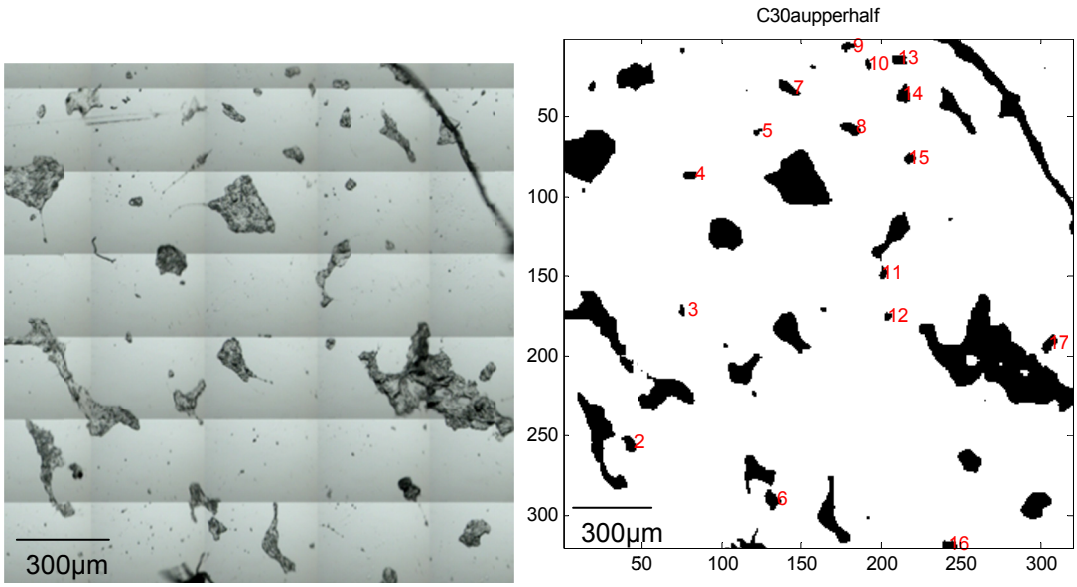
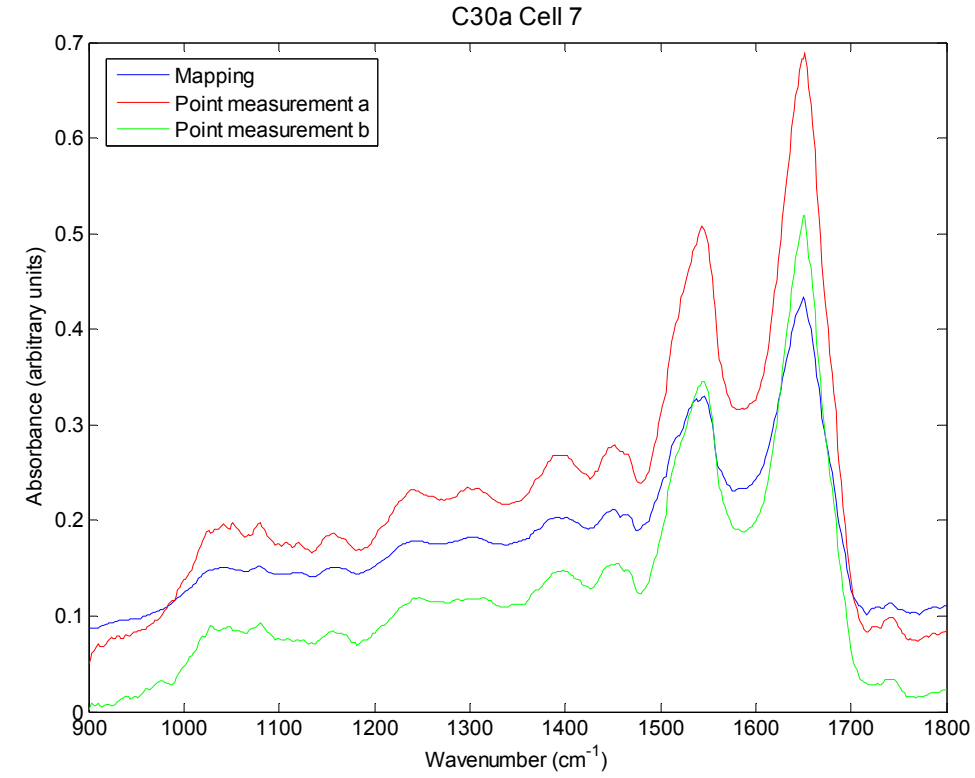
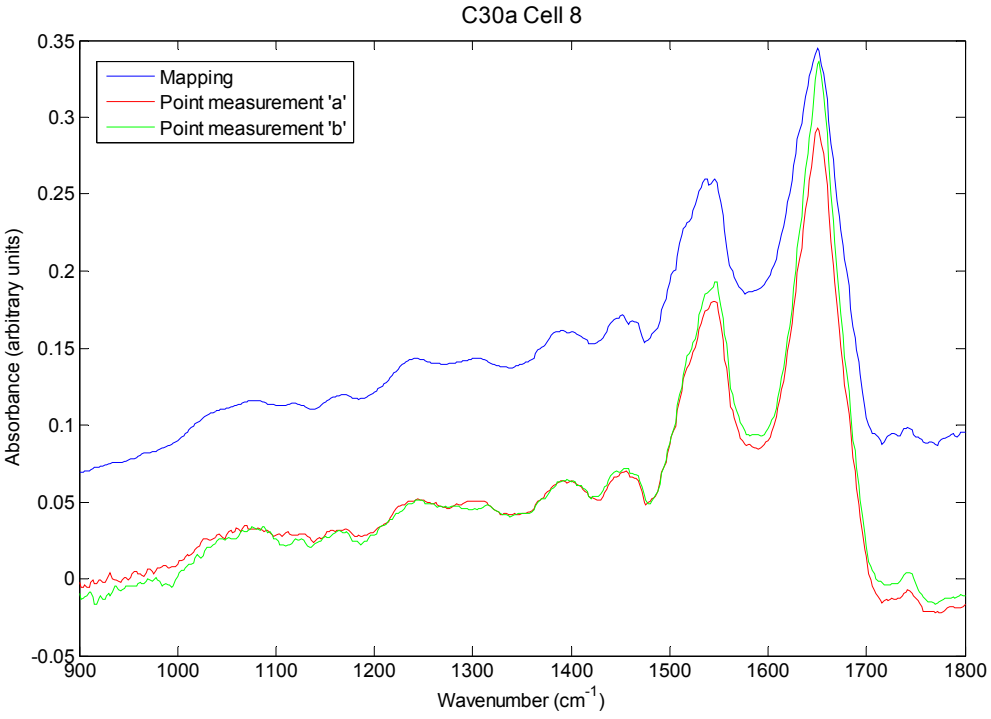


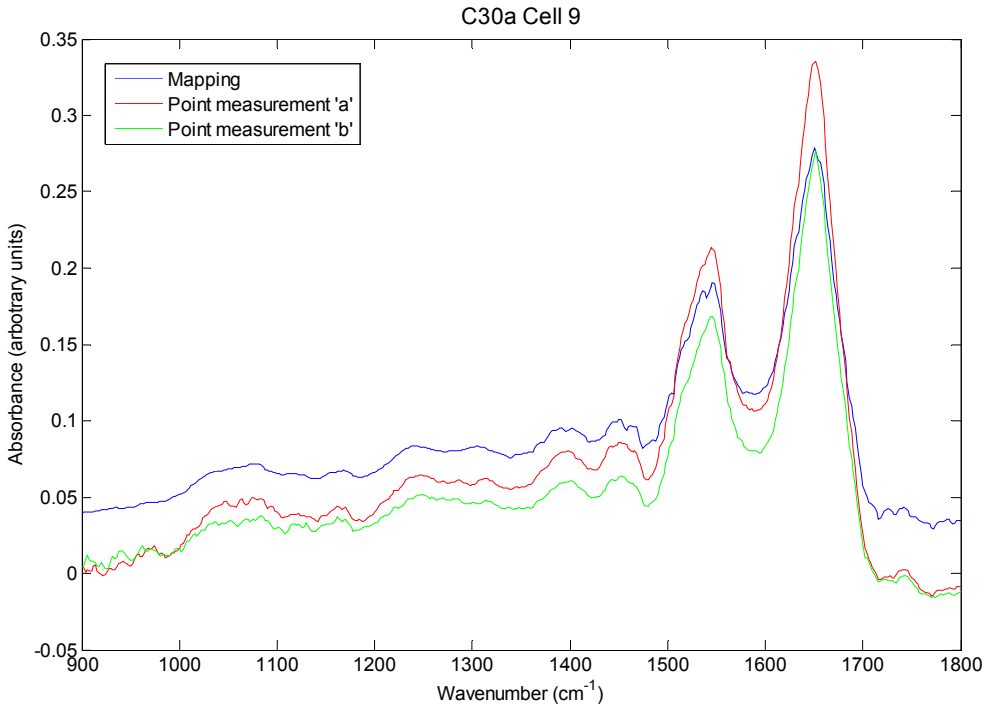
Figure 5-10 A) White light image of the unstained cell slide, B) Binary mask image of the same region.

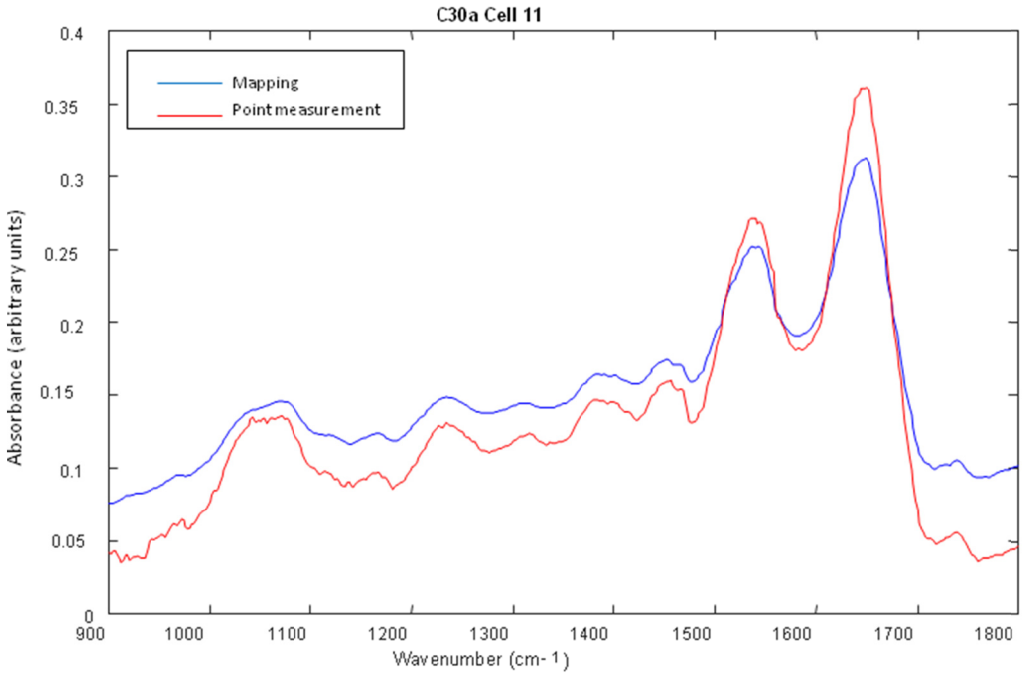
The white light and binary mask shows the presence of cells and debris on the slide. Although the use of the 1650cm^{-1} peak identifies large pieces of debris in addition to cells, the size filter on the binary mask does not label these large regions as cellular. However, smaller regions of debris may be misidentified as cells.

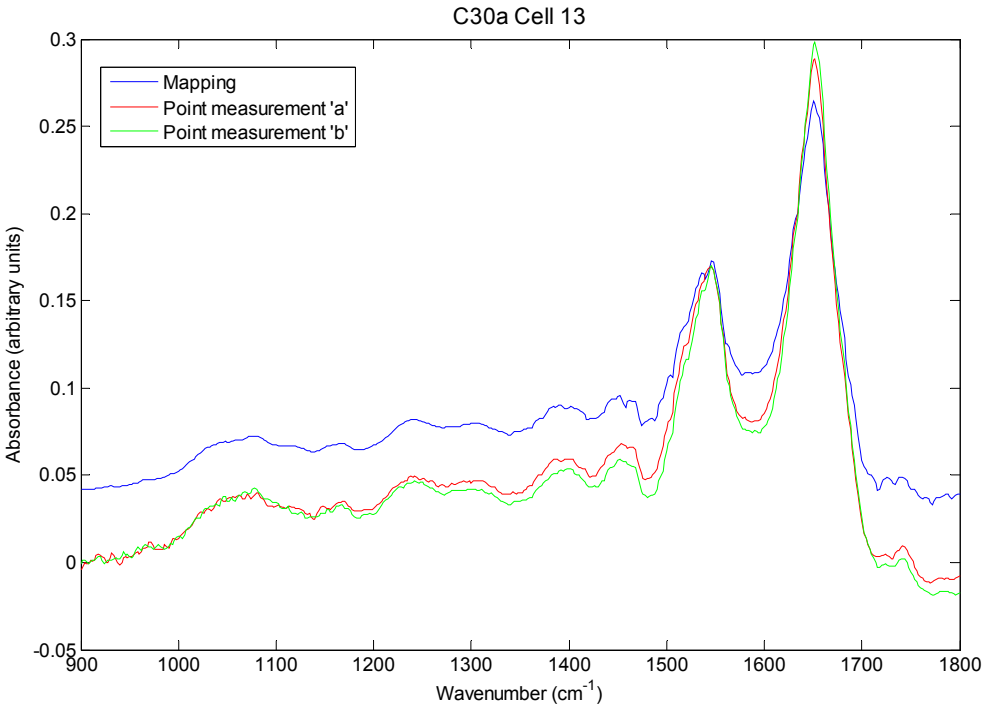
Point measurements were performed for 10 regions identified as cells by visual inspection of the unstained slide. All these regions were identified and labelled as cells by the binary mask. The mean spectrum for the cell from the binary mask and the point measurement spectrum from the same cell (or two spectra, labelled a and b, for larger cells) were compared. Selected cell spectra are plotted in Figure 5-11.

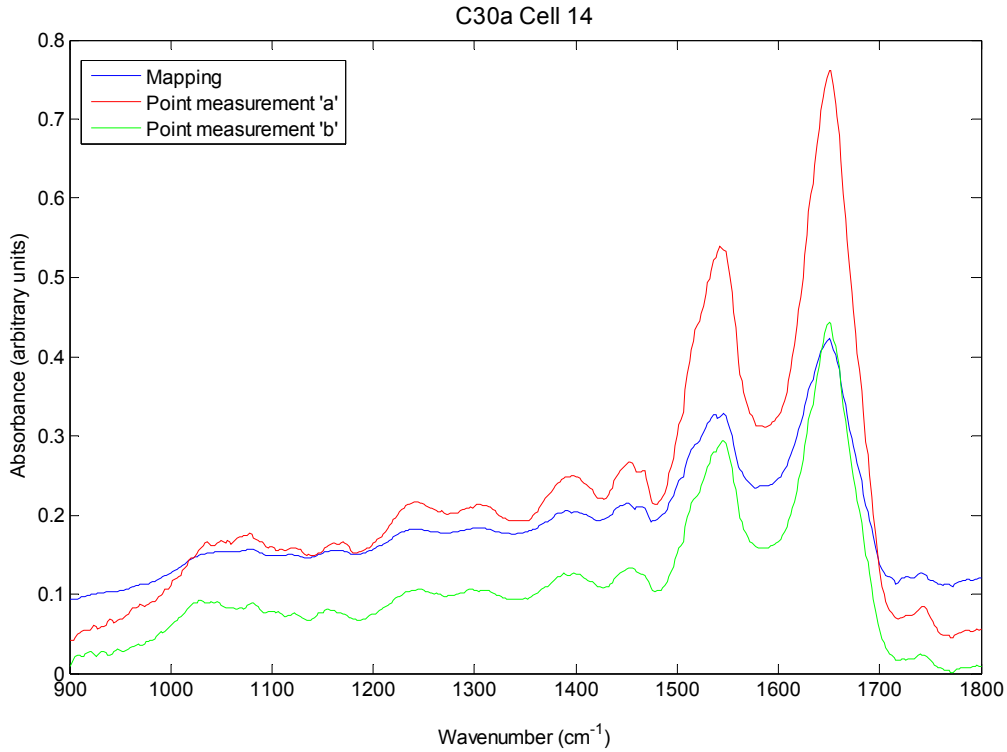

C30a Cell 7
Mapping
Point measurement a
Point measurement b
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
900
1000
1100
1200
1300
1400
1500
1600
1700
1800


C30a Cell 8
Mapping
Point measurement 'a'
Point measurement 'b'
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
900
1000
1100
1200
1300
1400
1500
1600
1700
1800


C30a Cell 9
Mapping
Point measurement 'a'
Point measurement 'b'
Absorbance (arbotrary units)
Wavenumber (cm-1)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
900
1000
1100
1200
1300
1400
1500
1600
1700
1800


C30a Cell 11
Mapping
Point measurement
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
900
1000
1100
1200
1300
1400
1500
1600
1700
1800


C30a Cell 13
Mapping
Point measurement 'a'
Point measurement 'b'
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
900
1000
1100
1200
1300
1400
1500
1600
1700
1800


C30a Cell 14
Mapping
Point measurement 'a'
Point measurement 'b'
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

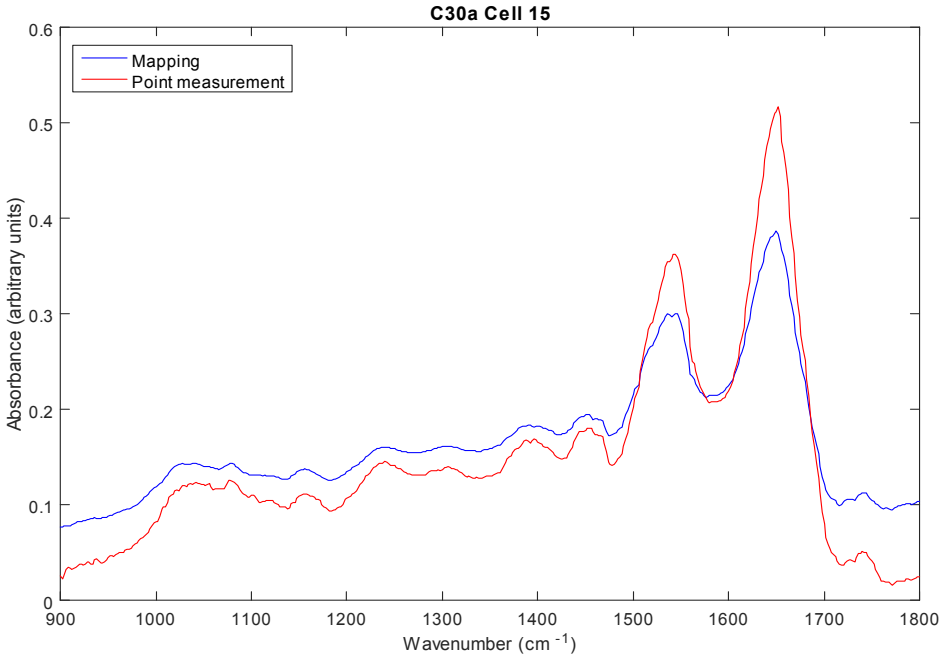


Figure 5-11 Comparison of binary mask cell spectra ('mapping') with point measurement spectra from the same cell. Cell labels correspond to the labels on the binary mask in Figure 5-10 above.

All the regions identified using the binary mask appear to contain biological spectra. Looking at each cell comparison individually, in every case the same peaks are present in both mapping and point measured spectra, strongly suggesting the spectra are from the same cell. In examples where there are clear differences between the cells e.g. at 1080cm^{-1} there is a small sharp peak in Cell 7 or Cell 15, but in Cell 11 there is a much broader peak across 1020-1090cm^{-1}, these features are consistently seen in the two modalities for the same cell.

The relative intensities are similar, although not identical: the mean spectrum from the mapping measurement is composed of all pixels within the cellular region, so will contain more spectra than the point measurements, and

consequently in most cases has a higher signal strength, and a greater smoothing effect from averaging a larger number of measurements.

On this evidence we concluded that the binary mask was successfully identifying cellular regions, extracting their spectra, correctly labelling them, and equivalent to point measurement of individual cells, and could potentially be used for automating cell detection.

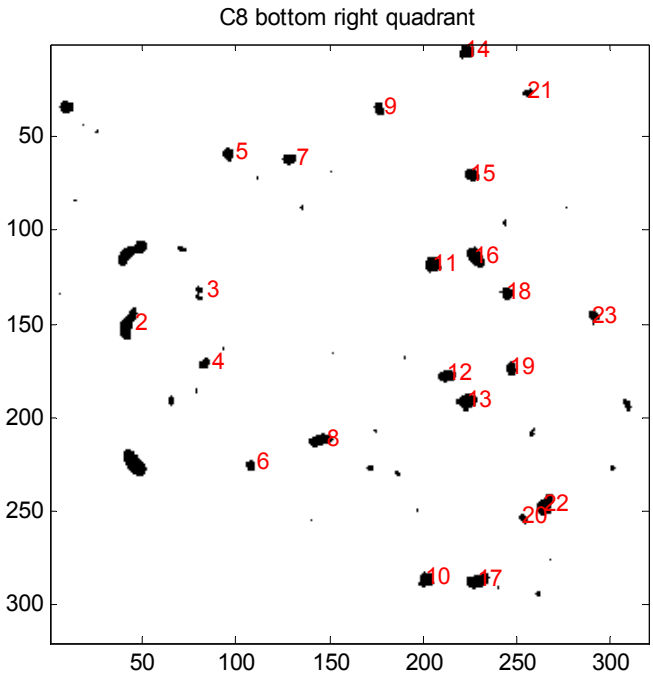
5.7.3. Atmospheric water vapour correction

Water vapour within the measurement chamber can cause artefacts in the spectral signal. There are a number of methods for minimising this effect. Spectrometer measurement chambers can be continuously purged with anhydrous gas to exclude water vapour from the measurement: this requires a continuous supply of gas whilst samples are measured and is therefore a relatively costly approach, and not appropriate for the current project. Alternatively the artefacts from the water vapour can be removed in post-processing. This involves subtraction from (or addition to) the spectrum to counteract the effects of the water spectrum on the measured sample spectrum, with the aim of removing all contributions of water vapour. We initially used a standard algorithm used by the spectrometer manufacturer, however early measurements showed an artefact suggestive of water vapour.

When inspecting the spectra from the initial cell maps, many spectra showed evidence of an artefact that created a rapidly oscillating spectral trace. Intriguingly, the artefact was present in only some spectra, but not others, even when measured from cells close together on the same slide. An example of this

is shown in Figure 5-12, which plots all cellular spectra from a single map. The rapidly oscillating artefact is seen in some spectra, but is absent in others.

A



B

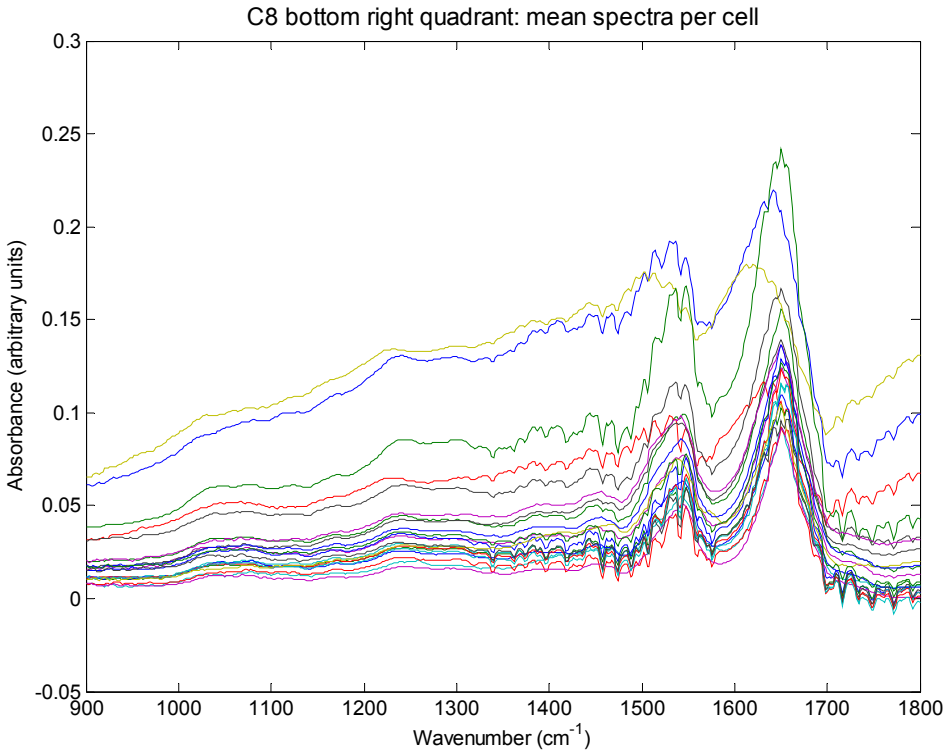


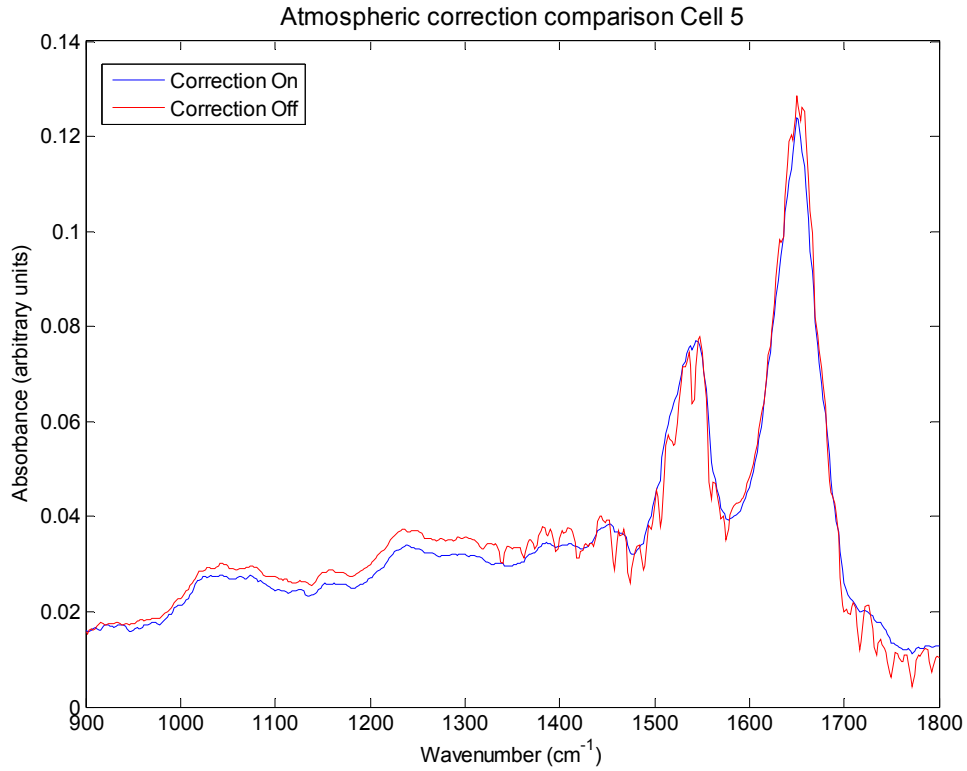
Figure 5-12 A) Binary mask showing cellular regions and B) raw spectra (mean spectrum per cellular region) from cell sample C8 bottom right quadrant.

This problem has been recognised previously for FTIR measurement as being caused by atmospheric water vapour. The spectral contribution of water vapour should be reduced by subtracting the background signal (we measured a blank area of the slide prior to each measurement, this is subtracted automatically on the Perkin Elmer), however algorithms to correct for any residual water vapour artefact have been developed (Bruun et al. 2006). The Perkin Elmer system automatically applies an in-built algorithm to the spectra it collects, so the 'raw' spectra that it presents have already been corrected for water vapour. However the presence of this oscillating artefact in our spectra suggested that the correction algorithm was not working well for our cell spectra.

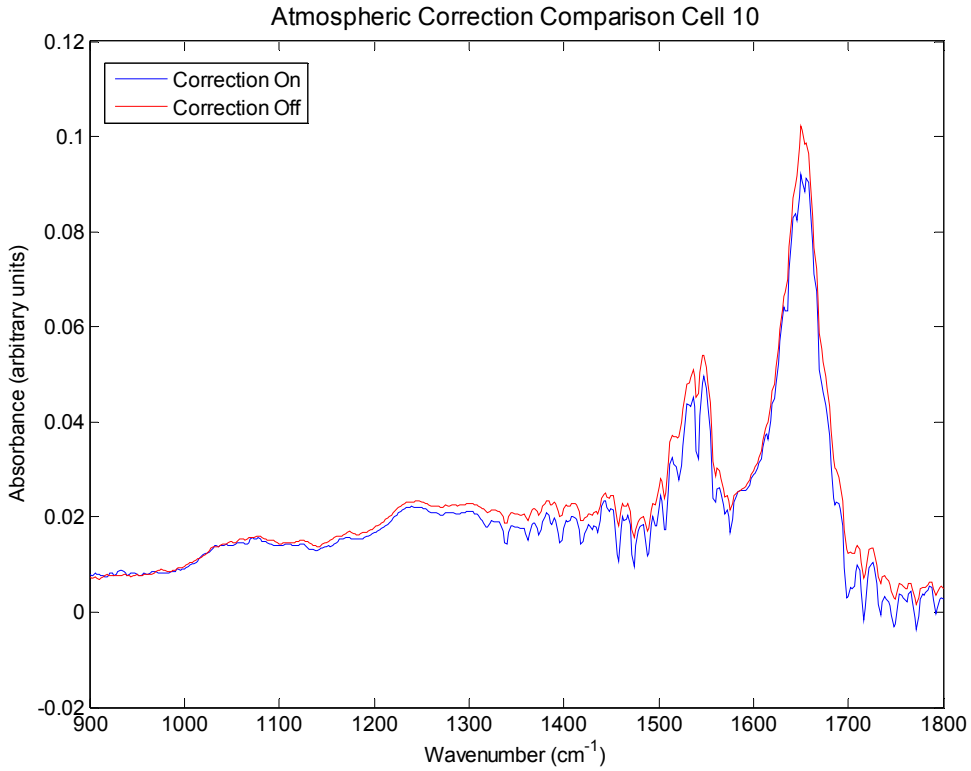
The correction algorithm used by Perkin Elmer is not publically available, and it is not possible to see the original spectra before and after applying this correction. However, it is possible to switch the automated correction off. We therefore compared the effect of the atmospheric correction on spectra by measuring the same region with the automated correction on, and measuring the spectra again with the correction off.

In order to compare the difference for individual cells, cells were selected from the mapped region to compare the differences between their mean spectra with atmospheric correction on or off. The spectra are shown in Figure 5-13.

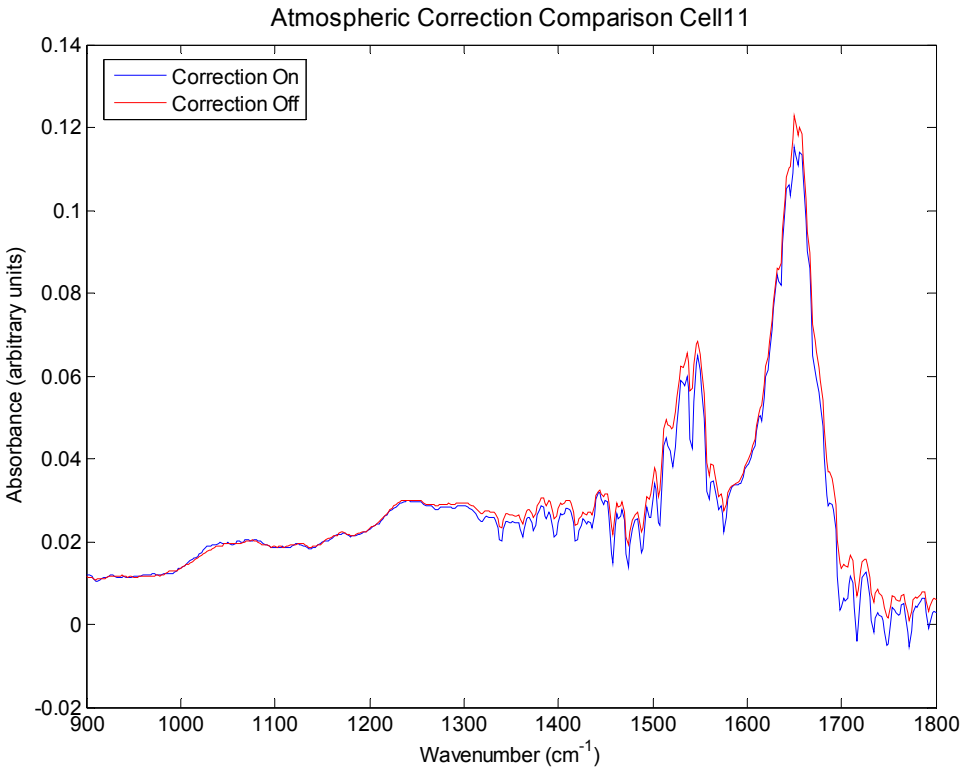
A


Atmospheric correction comparison Cell 5
Correction On
Correction Off
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

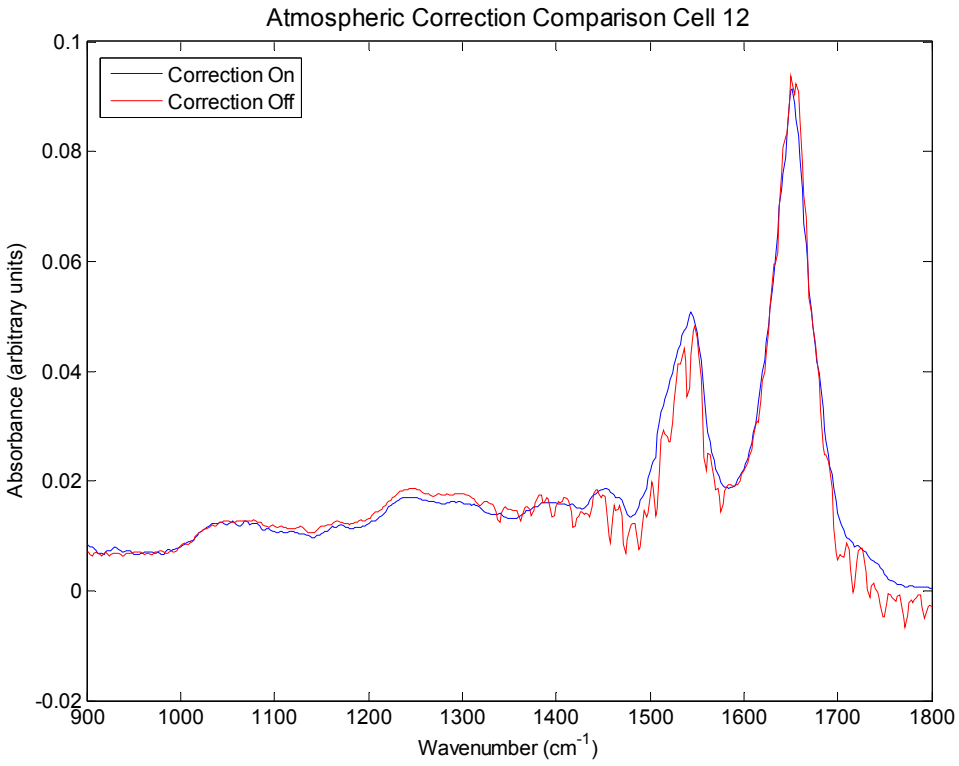
B


Atmospheric Correction Comparison Cell 10
Correction On
Correction Off
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

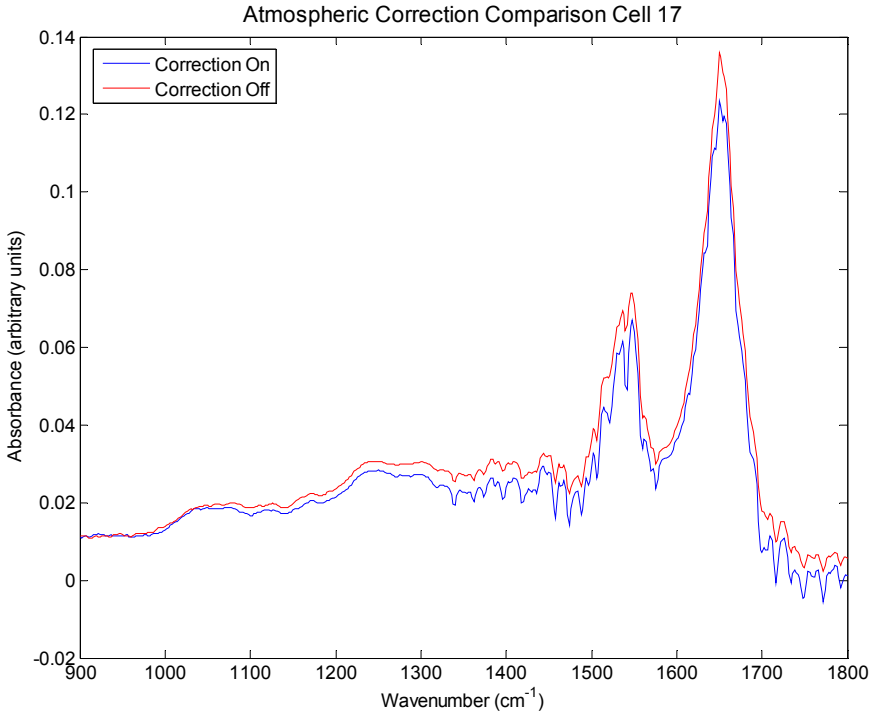
C


Atmospheric Correction Comparison Cell11
Correction On
Correction Off
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

D


Atmospheric Correction Comparison Cell 12
Correction On
Correction Off
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.1
0.08
0.06
0.04
0.02
0
-0.02
900
1000
1100
1200
1300
1400
1500
1600
1700
1800

E



F

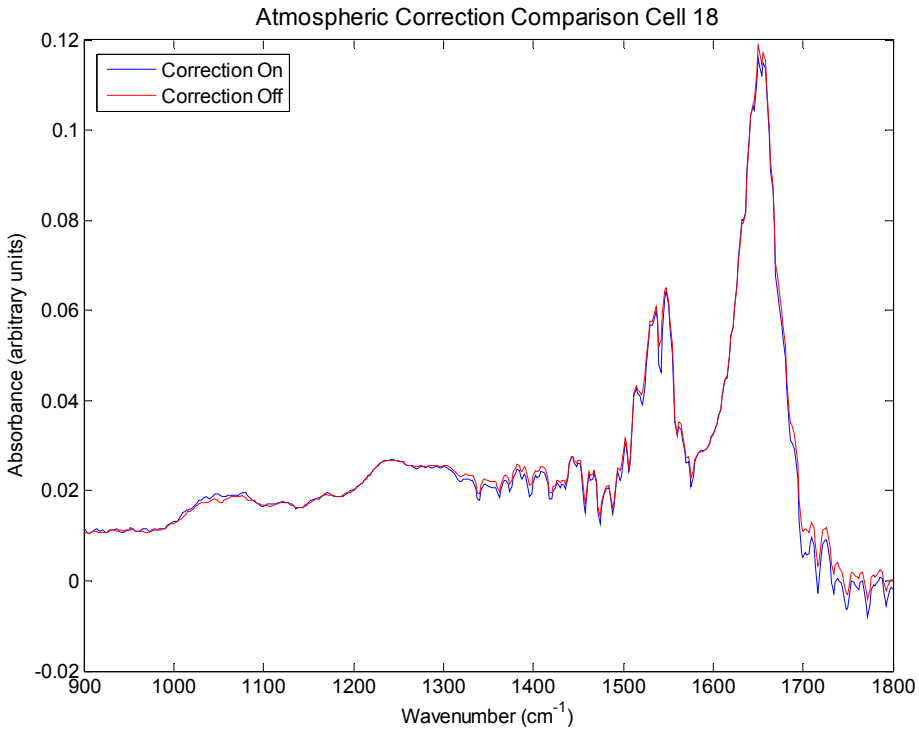


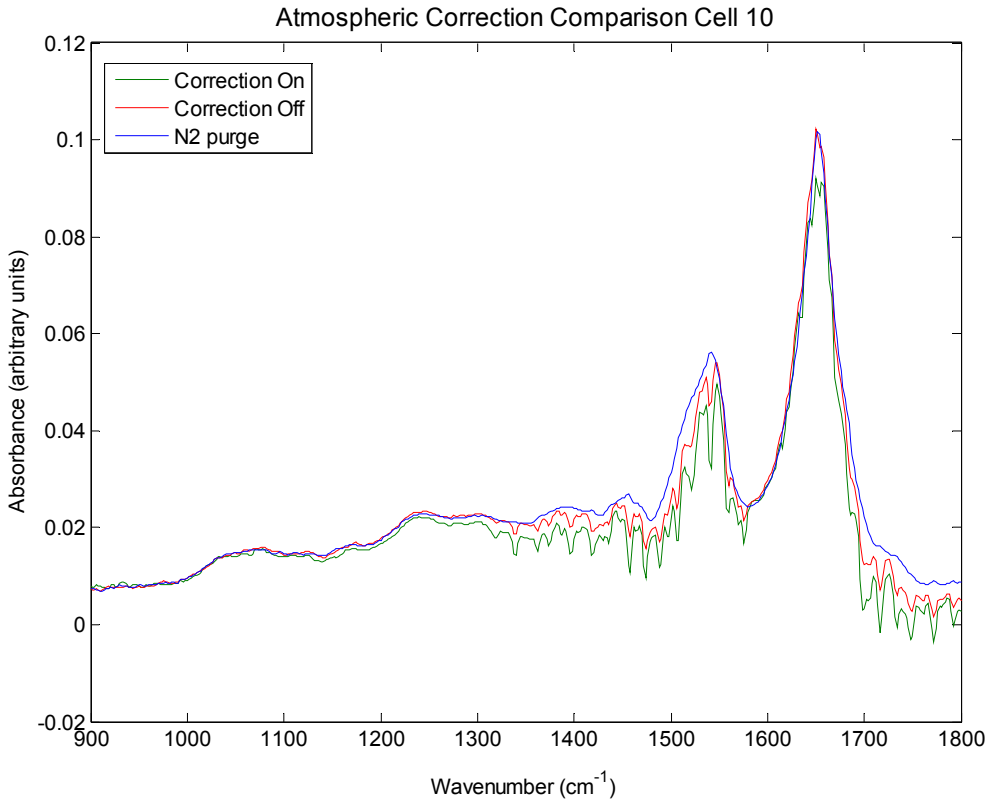
Figure 5-13 Comparison of effect of atmospheric correction for individual cells. Mean spectra per cell are shown with atmospheric correction on or off.

Looking at the spectra in Figure 5-13, the effect of the atmospheric correction is highly variable. In some cases, e.g. Cell 5 and Cell 12, the atmospheric correction appears to work well, removing the artefact from the spectrum measured with the correction off, and smoothing the spectrum. In some cases, e.g. Cell 11 or Cell 18, the correction algorithm has had very little effect. Whilst in others, such as Cell 10 or Cell 17, the algorithm appears to have introduced an increased artefact, greater than the small artefact that was present in the spectrum measured with no atmospheric correction.

This result suggested that water vapour, and problems with the PE algorithm to correct for it, were the cause of the artefact seen in our spectra. We therefore designed an experiment to test this, and develop our own algorithm to improve it.

To exclude atmospheric water vapour, the Perkin Elmer system can be connected to a cylinder of nitrogen gas, which then flows into the measurement chamber continuously to purge the system of air, and any concomitant water vapour. The measurement chamber is not completely airtight, but saturation with nitrogen gas allows measurement of a ‘pure’ signal with minimal contribution from water vapour.

To demonstrate the effect this has on our cellular spectra, a map was measured whilst purging the system with nitrogen using a high flow rate. The spectra from individual cells were compared against those measured with the atmospheric correction on or off, as shown in Figures 5-14 to 5-16.



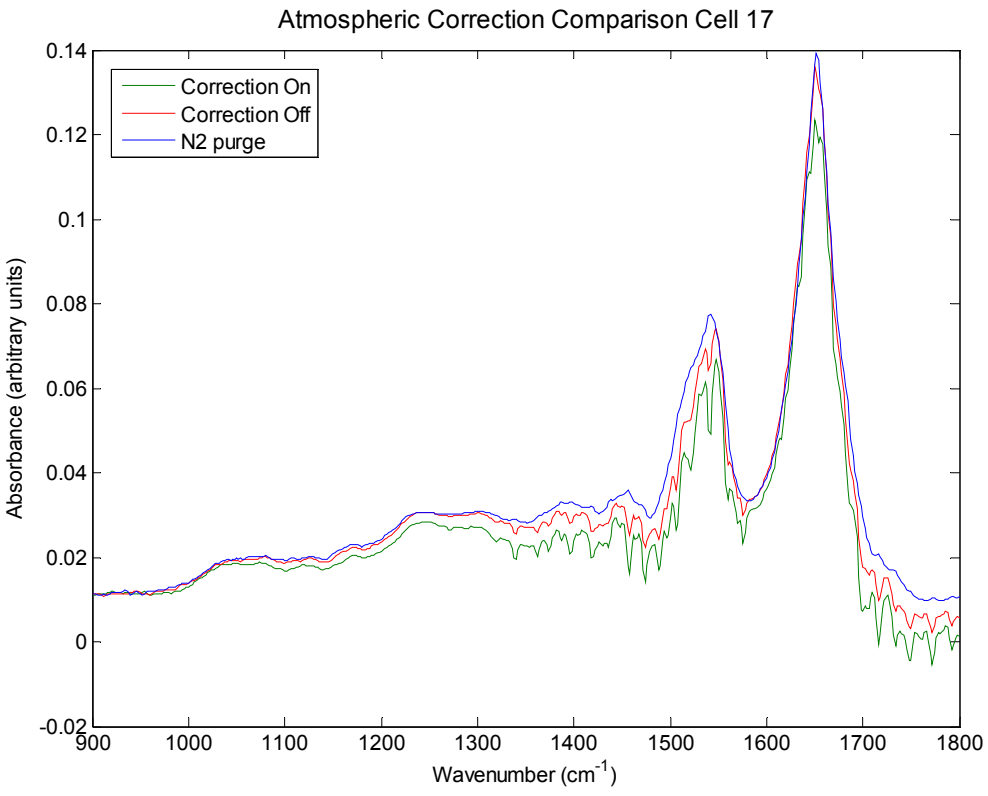


Figure 5-14 Comparison of measurements using nitrogen purge against those measured with normal atmospheric contribution. Each cell was measured with atmospheric correction switched on, and again with the correction switched off.

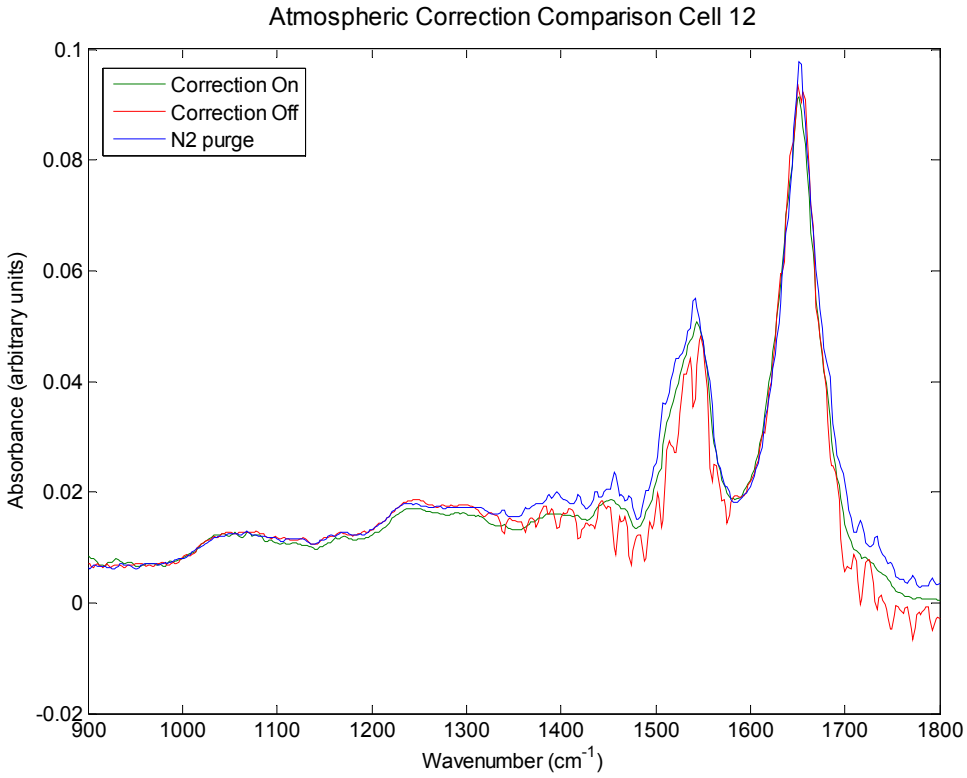


Figure 5-15 Cell measured with nitrogen running at half flow rate.

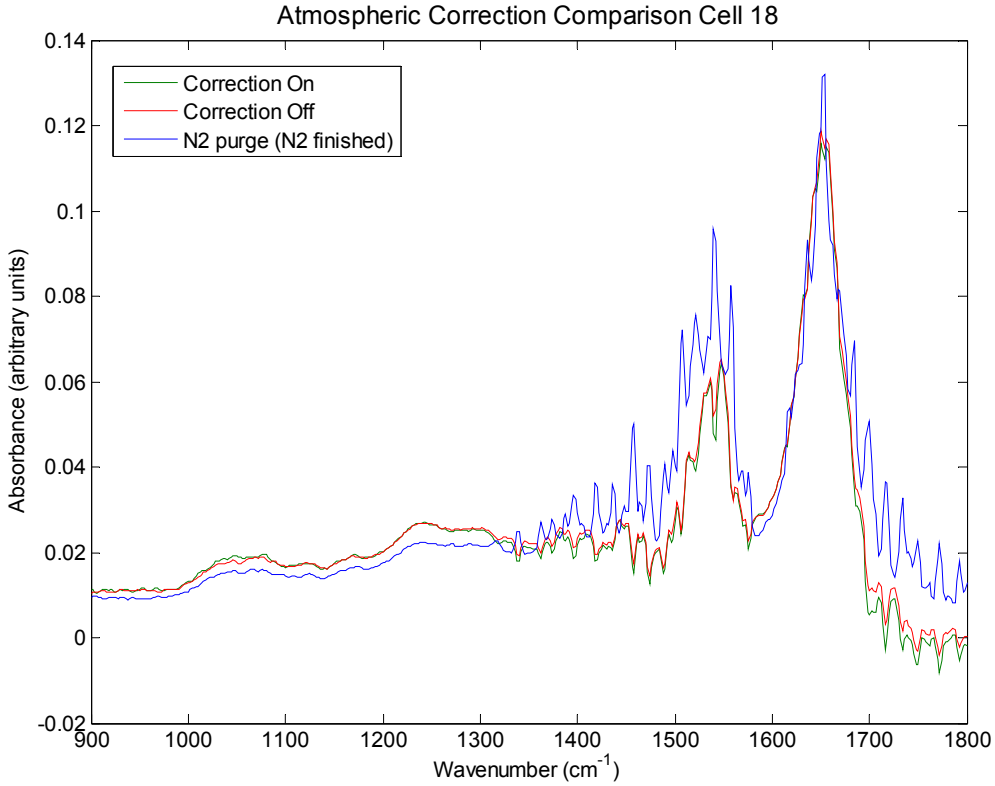


Figure 5-16 Cell measured 15 minutes after nitrogen purge stopped.

The effect of the nitrogen purge can be seen in Figures 5-14 to 5-16. When measurements were taken with the nitrogen purge, such as Cell 10 or Cell 17 in Figure 5-14, the spectral trace is smooth and free from artefact. Cell 12 (Figure 5-15) was measured with the flow rate halved, and some very minor artefact is seen in the spectrum. Further measurements were taken after the nitrogen was switched off, and Figure 5-16 shows the spectrum of Cell 18 at 15 minutes after stopping the nitrogen. Here there is a large artefact and the spectrum is highly distorted. This effect is greater than that seen in the 'normal' measurements because the background (which is automatically subtracted from the spectral signal) was measured with a purged system. Therefore, even though the water vapour content is likely to be the same across all 3 measurements, the water vapour relative to the background is much higher in the measurement performed after the nitrogen purge.

This confirmed that water vapour was the source of the artefact in our earlier measurements. To correct for this in future measurements, we developed an algorithm to subtract the spectral contribution from water vapour. In order to characterise the contribution from water vapour, a background spectrum without water vapour is needed for comparison.

The water vapour is seen in the spectrum from around 1300-2000cm^{-1}, and again at around 3500-4000cm^{-1}. There is a large peak around 2350cm^{-1} due to atmospheric carbon dioxide. The nitrogen purge reduces both the water vapour and CO_2 signal but does not eliminate them. In this experiment there was no further change in signal after 5 minutes of purging the measurement chamber with nitrogen, suggesting that this was the greatest level of saturation with nitrogen we could achieve.

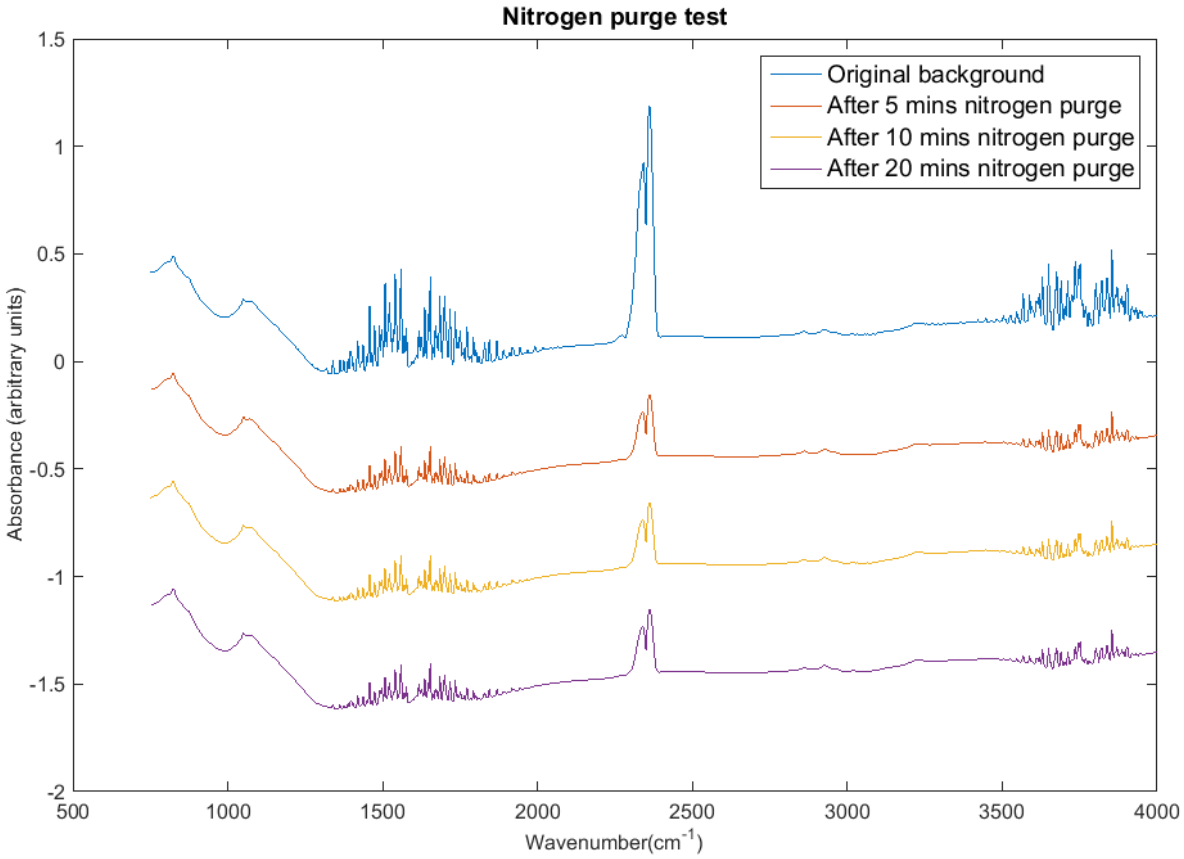


Figure 5-17 Comparison of original background spectrum with spectra obtained after purging the Perkin Elmer measurement chamber with nitrogen gas. Spectra are offset for clarity.

This allows a ratio to be calculated of the original background (with water vapour) versus the background after nitrogen purge – in effect an 'atmospheric' spectrum, containing the spectral contribution from the atmosphere. This can then be used to subtract from the sample spectrum to reduce the contribution of water vapour from our sample spectra (the CO_2 peak would also be subtracted but this occurs outside the fingerprint wavenumber range used for analysis).

Figure 5-18 shows the mean spectra from C8 bottom right quadrant after the algorithm has been applied (see raw spectra in Figure 5-12 above for comparison). The water vapour artefact has been almost completely removed,

and there is very little noise in the spectra. Consequently we decided to measure all samples with the PE atmospheric correction switched off, and use our own algorithm.

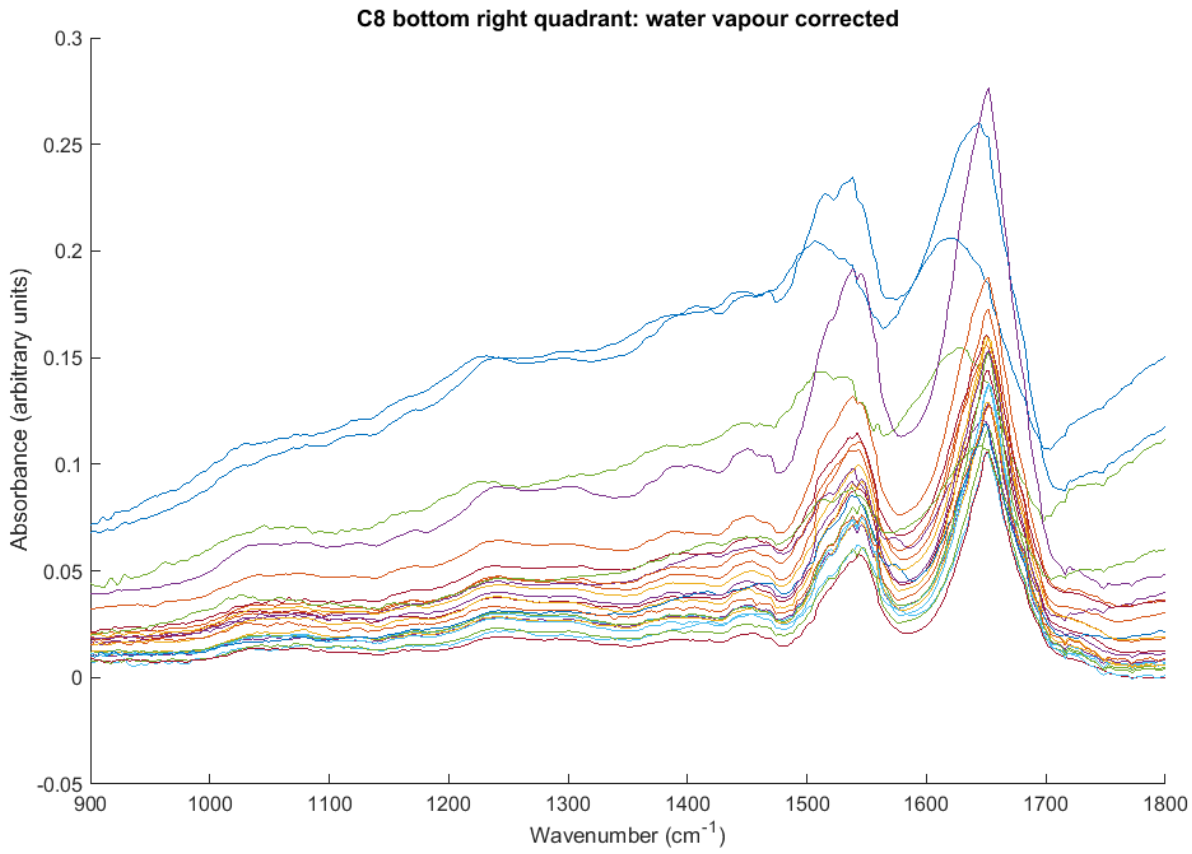


Figure 5-18 C8 bottom right quadrant, mean spectra per cell after water vapour correction.

5.8. Summary of measured samples

From previous work within the Biophotonics unit, sample sizes of around 100 are sufficient to capture a high degree of within-group variation. We therefore estimated that a sample size of 100 cells in each pathology group would be sufficient to inform a robust training model. To ensure that this captured

sufficient between-patient variability, again based on previous work within the research group, we estimated that 20 patients would be needed for each pathology group. We therefore required an average of 5 representative cells per slide to meet our target of 100 cells in each group. Because of the potential for low numbers of dysplastic cells amongst dysplasia/adenocarcinoma samples, we aimed to include 25 patients in this group, and 20 in each of the other groups.

Cell brushing samples were only included if the histology from that region was certain. Discrepancies were possible because brushing occurs over a larger area than the size of a biopsy, and there were a number of possibilities where the samples could be discordant. For example, if a brushing was taken from a region that appeared dysplastic, with dysplasia on other biopsies in the oesophagus, but not seen in the biopsy taken from the region of brushing, the sample was excluded. Similarly, if a brushing was taken from a region thought to be non-dysplastic Barrett's, with no dysplasia on the biopsy from the region of brushing, but dysplasia was found on other biopsies taken at the same endoscopy, the sample was excluded.

A total of 107 patients were recruited to the study. Of these, 69 patients were included in the study, contributing a total of 91 samples. Each patient only contributed one type of oesophageal sample, but stomach samples were collected almost entirely from patients who also contributed oesophageal samples. We achieved our recruitment target for every pathology group.

Table 5-1 Summary of measured data

	NSQ	BE	DYS/AC	STOMACH	TOTAL
Patients	20	21	26	21	69*
No. of brushings	21	22	27	21	91
FTIR maps	22	34	44	22	122
Total spectra	3,891,600	5,222,400	7,270,400	4,300,800	20,685,200
Cell regions identified	786	2074	3164	1041	7,065
Total spectra in cell regions	32,807	63,640	94,876	37,693	229,016

NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/adenocarcinoma. *Stomach samples were taken largely from patients who also contributed an oesophageal sample.

5.9. Developing a training dataset

The prepared cell slides contained a mixture of different cell types, along with cellular and non-cellular debris. In order to inform a training dataset accurately, we aimed to include only those cells that were representative of the appropriate pathology type. We planned to achieve this by staining the slides after FTIR measurement, recording microscope images of a representative number of stained cells, correlating the position of the photographed cells with their corresponding cell number on the binary mask from the FTIR map of that

particular sample, then showing the images to 2 cytopathologists for classification. This process is illustrated in Figure 5-19.

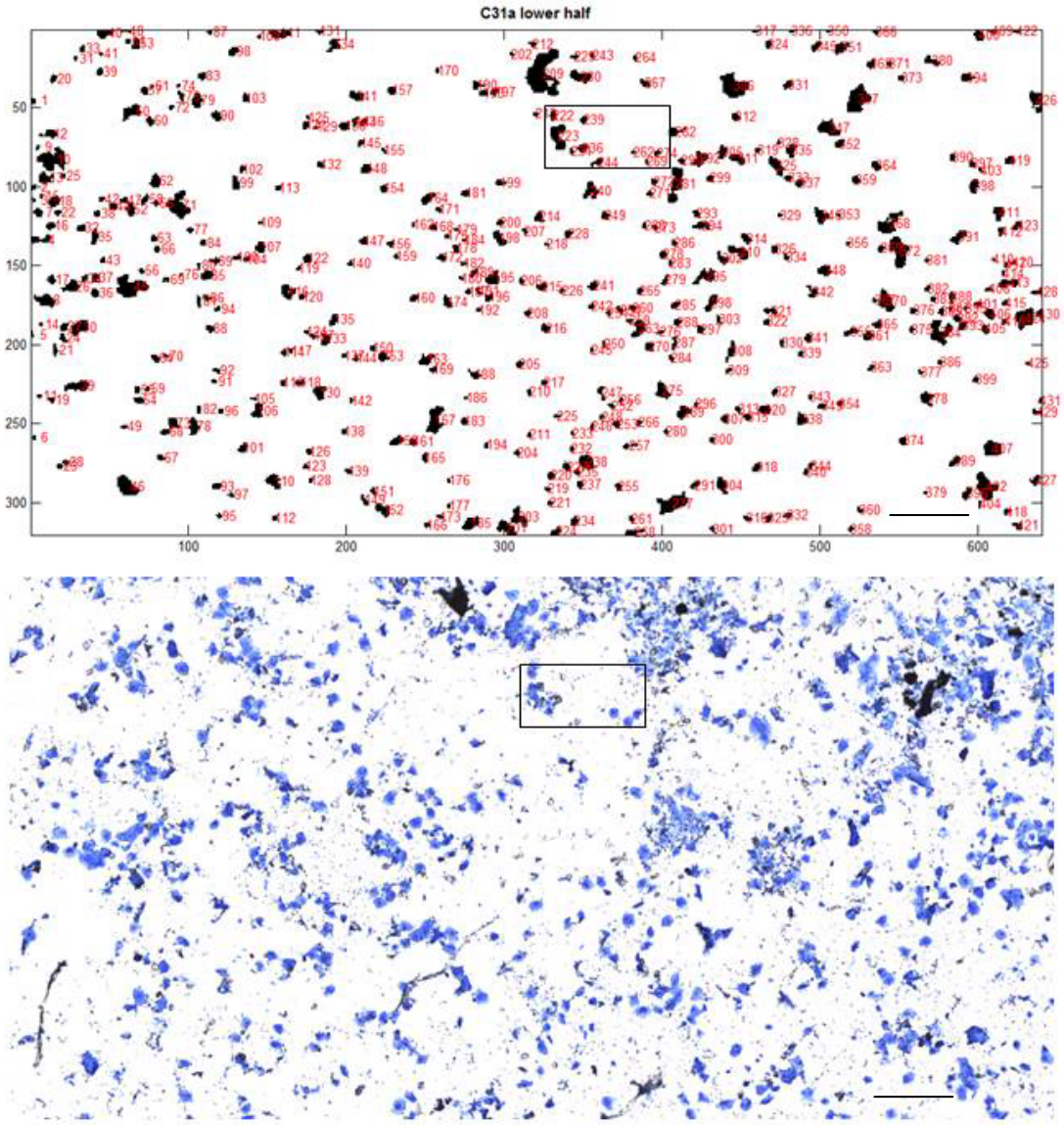


Figure 5-19A: Developing a training dataset. The whole measured region: binary mask (upper image, with each cell numbered) and corresponding stained cell slide (lower image). With high cell density, identifying which cell on the slide corresponds to the same cell on the binary mask could be very challenging. In both images the bar represents 300µm. The highlighted region is enlarged in Figure 5.19B.

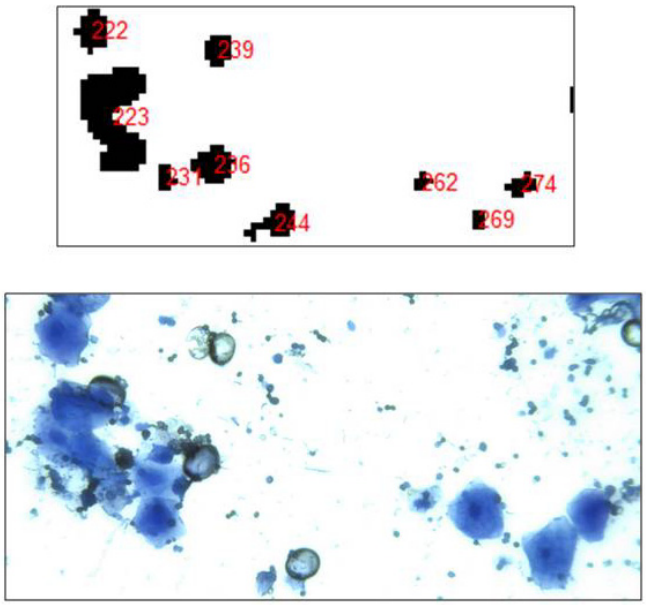


Figure 5-19B Enlarged region of binary mask highlighted in 5-19A, and corresponding region of slide. This enabled labelling of the individual cells. The correspondence between the stained slide and the binary mask was often imperfect. Only cells that could confidently be identified as a particular cell in the binary mask were included in the training dataset.

5.9.1. Cell staining

After infrared measurement samples were stained using a 'Diff-Quik' staining technique. Staining was performed by dipping the cell slides into Diff-Quik fixative reagent (methanol based fixative, Thermo Scientific) for 10 seconds, followed by Diff-Quik solution I (an eosinophilic red stain, to identify cytoplasm, Thermo Scientific) for 10 seconds, then Diff-Quik solution II (a basophilic blue stain, to identify nuclear material, Thermo Scientific) for 10 seconds, then rinsing with a buffer solution at pH7.2. The slides were then air dried and transferred for storage.

5.9.2. Photographing cells and correlating with binary mask

We had initially hoped to be able to photograph cell slides automatically with a slide scanner, but unfortunately this was not possible. The CaF_2 slides were not compatible with the available slide holders in the two slide scanners we tried (University College London), which required a slide of standard rectangular dimensions and thickness. We attempted mounting a CaF_2 slide onto a regular slide, which overcame the problem with the slide holders. Unfortunately, however, by changing the distance of the slide surface from the lens, one of the slide scanners accepted the slide but produced an out of focus image, and the other would not process the slide as it recognised that the surface was less than a minimum required distance from the lens, and risked scratching the lens.

As a result, I manually reviewed and photographed cells on the slides. This process involved identifying individual cells on the stained slide and correlating the position of the cell with the number assigned to it on the binary mask. The cell was photographed, and labelled with the number it had been assigned on the binary mask – this would then allow us to extract the spectra from the correct cell into the training dataset after cytopathology review of the cell photograph.

The process of manually correlating the position of the cell on the slide with the binary mask was very time consuming. The binary mask worked variably well in identifying every cell – on some slides there were many regions of cellular debris which were labelled as cells (see Figure 5-20). This was more readily apparent on the stained slides, and in some cases there was a great deal of debris and very few cells indeed. In other cases, structures that appeared cellular were not picked out by the binary mask, though this occurred

infrequently. There were also a few slides where some cells were ‘missing’ and had presumably been washed off the slide by staining.

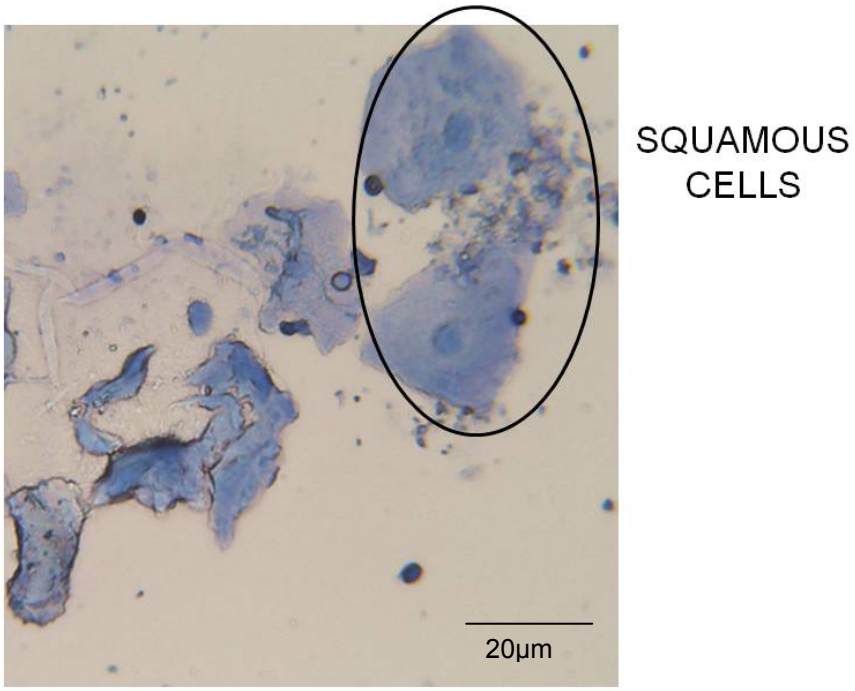


Figure 5-20 Example of cell debris/degenerative cells on the slide. There are two squamous cells (labelled), but the other material is cell debris or possibly degenerative squamous cells. Such material was often identified by the binary mask as being cellular, but was not included in the training dataset.

Whilst doing the photographing it also became apparent that there was, on average, a difference between the cellularity of the different pathology groups on the slides. The normal squamous samples tended to have much lower cell density than Barrett’s samples, and the dysplasia/adenocarcinoma samples were even more heavily cellular, and showed greater density of red blood cells.

It was clear from initial review of the slides that there would be far more normal squamous cells identified to include in the training model than Barrett's, dysplasia/adenocarcinoma, or stomach cells. Because of the time taken to identify and photograph the cells, we decided to include only half of the normal squamous samples in the training dataset, since this would be sufficient to include more cells than the other groups, and a balanced training dataset is required to build an accurate classification model. Therefore 10 normal squamous samples, and all of the other samples (22 Barrett's, 13 dysplasia, 14 adenocarcinoma, 21 stomach) were reviewed and photographed. For the normal squamous samples, a maximum of 30 cells per sample were selected to be photographed for review. For the Barrett's, dysplasia/adenocarcinoma and stomach samples, it was clear that mixed cell types were present on the slides, with many cells of squamous morphology. Every cell region that was considered to possibly be a glandular cell was photographed, along with some examples of cells that were clearly of squamous morphology.

This step in the development of the training model proved to be very challenging, as it became apparent that much of what was on the slides was unidentifiable debris, or poorly preserved cells. This drastically limited the number of cells that could potentially form part of the training model. It also became clear that the overwhelming majority of cells on the Barrett's, dysplasia and adenocarcinoma slides appeared squamous in morphology, further reducing the number of cells to be included in the training model. This proved to be one of the major limiting factors in this study. These problems, and potential improvements for future work, are discussed further in the discussion, section 5.14.

5.9.3. Cytology review

Individual cell photographs were then presented to 2 cytopathologists, who reviewed each image together (not independently) to achieve a consensus on each image (if no consensus was achieved the cell was not included in the training model). The pathologists were blinded to the endoscopy and biopsy results for the sample from which the cell was taken.

Cells were classified into one of five labels: ‘squamous’, ‘glandular’ (considered to be non-dysplastic Barrett’s if from a Barrett’s sample, or stomach if collected from stomach fundus), ‘dysplastic’ (i.e. dysplasia/adenocarcinoma), ‘degenerative cell’, or ‘cell debris/other’.

5.10. The training dataset

As mentioned above, the number of cells included in the training dataset was greatly reduced by the finding of cellular debris and poor preservation of cells on the slides.

In total 6,008 spectra were included in the training set from 146 cell regions with consensus cytopathology (Table 5-2), from 20 patients. The term ‘cell regions’ is used because many of the regions contained multiple adjacent cells, which formed a single continuous cellular region identified on the binary mask. For brevity, the terms ‘cell’ and ‘cell regions’ will be used interchangeably below.

The numbers of cells and patients in each group except normal squamous were far smaller than our desired sample size. This was particularly true for the stomach cells, from which only 2 cell regions were classified as glandular stomach cells.

Clearly this result was very disappointing, and is a major limiting factor on the ability of the training dataset to develop a robust classification model.

Table 5-2 Summary of numbers of patients, cells and spectra with consensus cytopathology included in training dataset.

	No. of patients	No. of samples	No. of cell regions	No. of spectra for analysis
Normal Squamous	8	8	107	4,780
Barrett's oesophagus	5	5	15	291
Dysplasia/ adenocarcinoma	5	5	22	735
Stomach	2	2	2	202
TOTAL	20	20	146	6,008

5.11. Classification performance of the training dataset

The disproportionate numbers of cells in each group creates a problem for building classification models: if there is a significant imbalance in the number of spectra per group, the model will tend to favour placing spectra in the largest group. To account for this, rather than using all 107 normal squamous cells, 22 normal squamous cells (the same number as the largest group, dysplasia/adenocarcinoma cells) were randomly chosen to build each classification model. Because the choice of which cells were included could affect the model performance, 50 iterations were run for each model, randomly

choosing 22 normal squamous cells each time. This allows a mean sensitivity and specificity to be calculated, along with a standard deviation.

With only 2 cells in the stomach group we decided not to include this in the training model, since it would be impossible to draw any firm conclusions from only 2 datapoints.

One of the Barrett's samples (C31A) had a markedly different appearance from every other sample measured (including all pathologies). In every spectrum of the sample there was a strongly negative peak at around 1180cm^{-1}. This could not be accounted for by the background spectrum, which was identical to the others. It is not clear what caused this unusual appearance in this sample, but as a result it was not included in the training set (the sample was still included in the test dataset when using whole sample data).

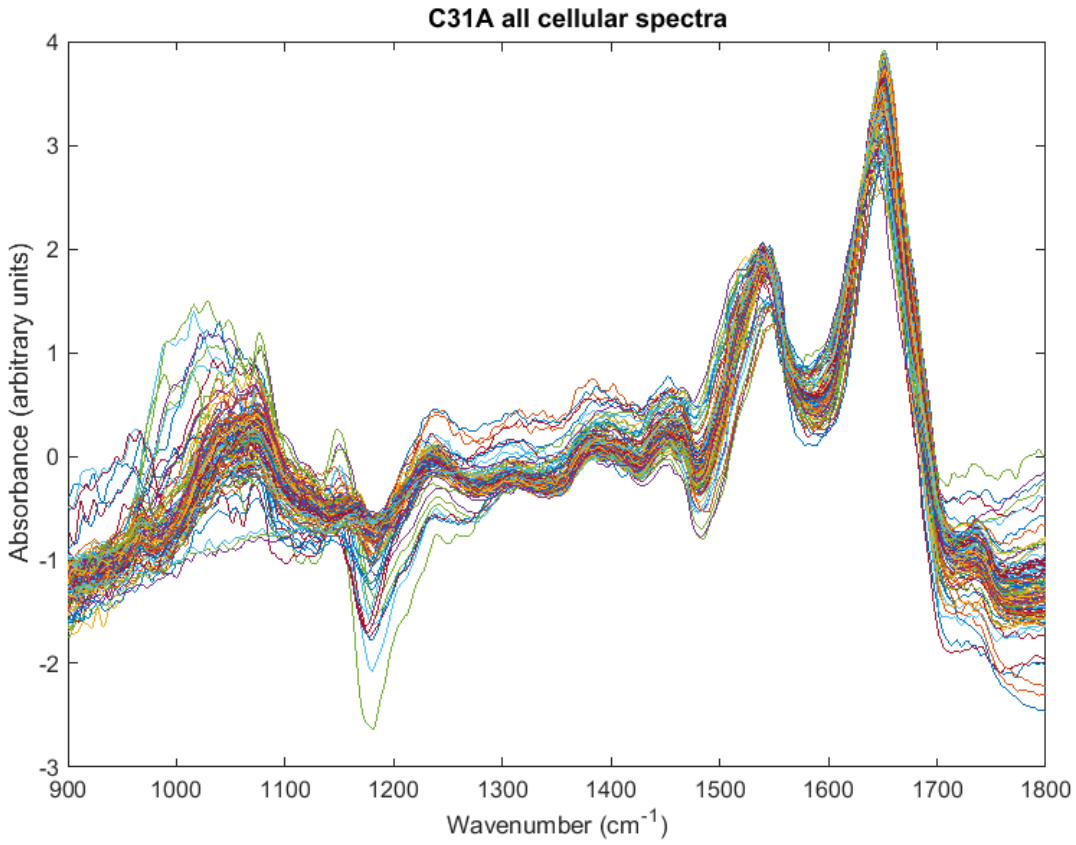


Figure 5-21 Barrett's sample C31A: all cellular spectra

Exclusion of this sample, which contributed 3 cells to the Barrett's training group, left only 12 cells remaining and significantly reduced the number of spectra in the Barrett's group.

Exclusion of outlying spectra was performed on each iteration of the training model. The measure of least squares distance from the reference spectrum (weighted mean spectrum) was calculated, and a histogram constructed to plot these values. Figure 5-22 shows an example histogram from the 3 group model below. The histogram shows a peak at around 0.75, and then 2 smaller peaks at around 1.5 and 2. This may represent the distribution of 3 distinct pathology groups. Spectra falling outside the range 0-2.5 were removed. This resulted in less than 5% of spectra being removed for each iteration.

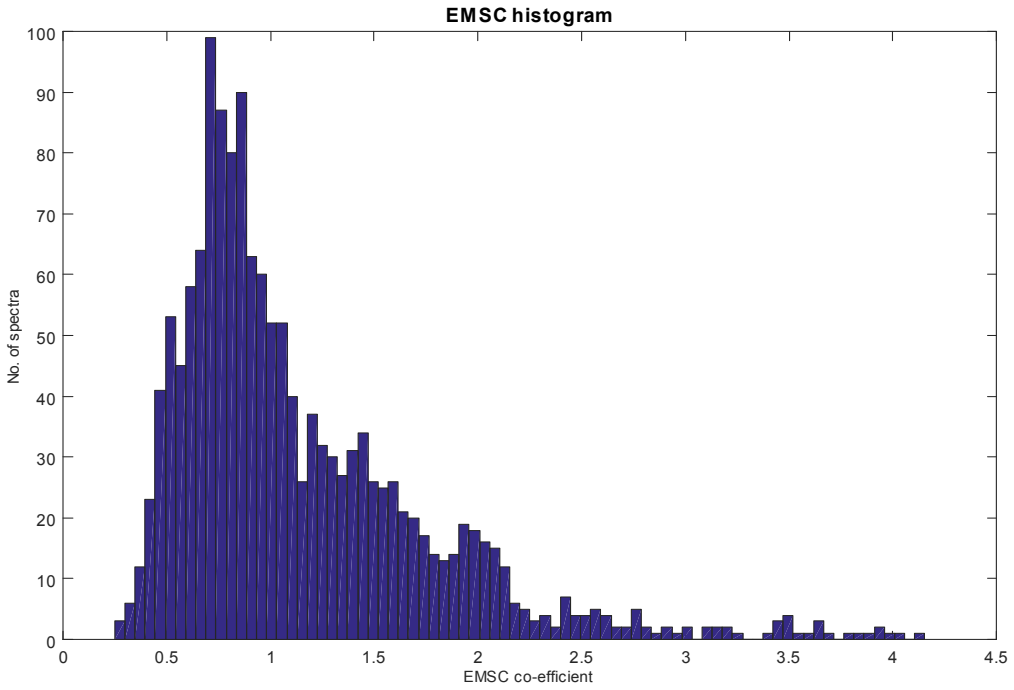


Figure 5-22 Histogram of least squares distance for each spectrum relative to reference spectrum for 3 group model.

Following exclusion of the outlying spectra on EMSC, the total number of spectra available for analysis in the training dataset is shown in Table 5-3.

Table 5-3 Summary of data included in the training dataset, following exclusion of sample C31A from the Barrett's group, EMSC exclusion of outliers, and random selection of 22 normal squamous cells (the exact number of normal squamous spectra will vary with each iteration, so the number shown is the first iteration chosen as an example).

	No. of patients	No. of cell regions	No. of spectra for analysis
Normal Squamous	8	22	726
Barrett's oesophagus	4	12	76
Dysplasia/ adenocarcinoma	5	22	678
TOTAL	17	56	1,480

5.11.1. 3 group model

The aim of this project was to develop a test that could be used either as a screening tool to identify Barrett's oesophagus, or as a method of Barrett's surveillance to detect dysplasia or adenocarcinoma. To achieve this clinical aim, the test must discriminate between normal, Barrett's and dysplastic samples.

Therefore the initial analysis compared the normal squamous, Barrett's and dysplasia/adenocarcinoma groups.

Prinicipal component analysis (PCA) was used to identify variance in the data: the 3 most discriminatory principal components (PCs) are plotted in Figure 5-23. There is reasonably good grouping for the spectra from each of the pathology groups. PC 7 separates the Barrett's spectra from the other groups, and PC 4 provides discrimination between the normal squamous and dysplasia/adenocarcinoma groups, although there is a degree of overlap. There is a small group of outliers from the normal squamous group, a few scattered outliers from the dysplasia/adenocarcinoma group, and a small subset of the Barrett's datapoints which lie much closer to the other two groups.

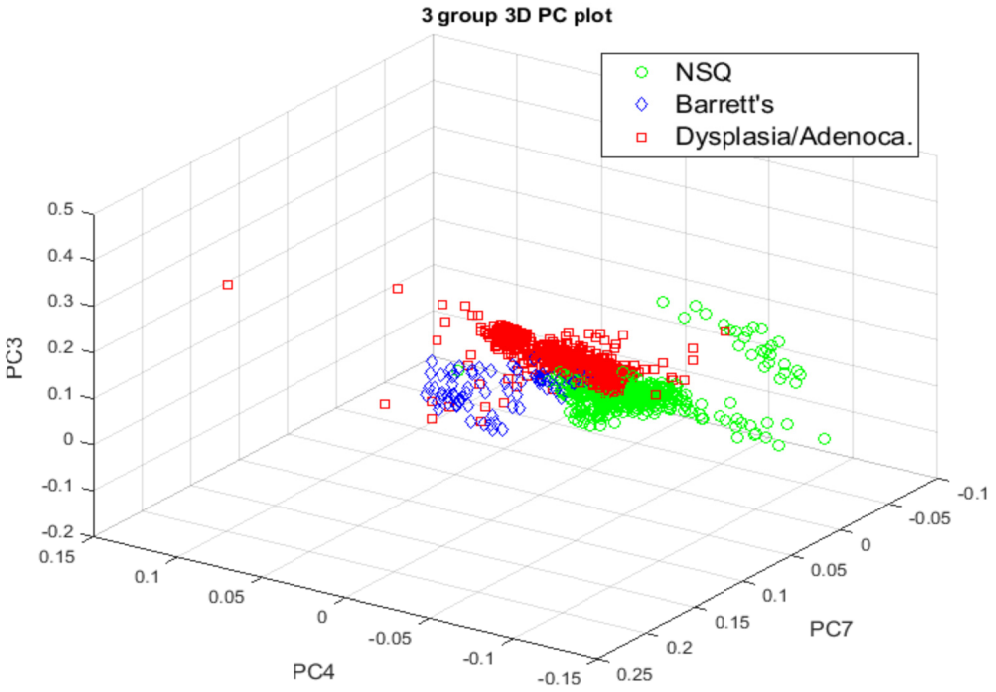


Figure 5-23 Scatter plot of all spectra, plotted by principal component (PC) score for 3 most discriminatory PCs: PC3, PC4 and PC7. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

The linear discriminant functions are shown in Figure 5-24. LDA achieves good grouping and separation of each of the pathology groups, but there remains a cluster of Barrett's datapoints close to the other two groups. Every one of these outlying datapoints was from a single sample, C64B.

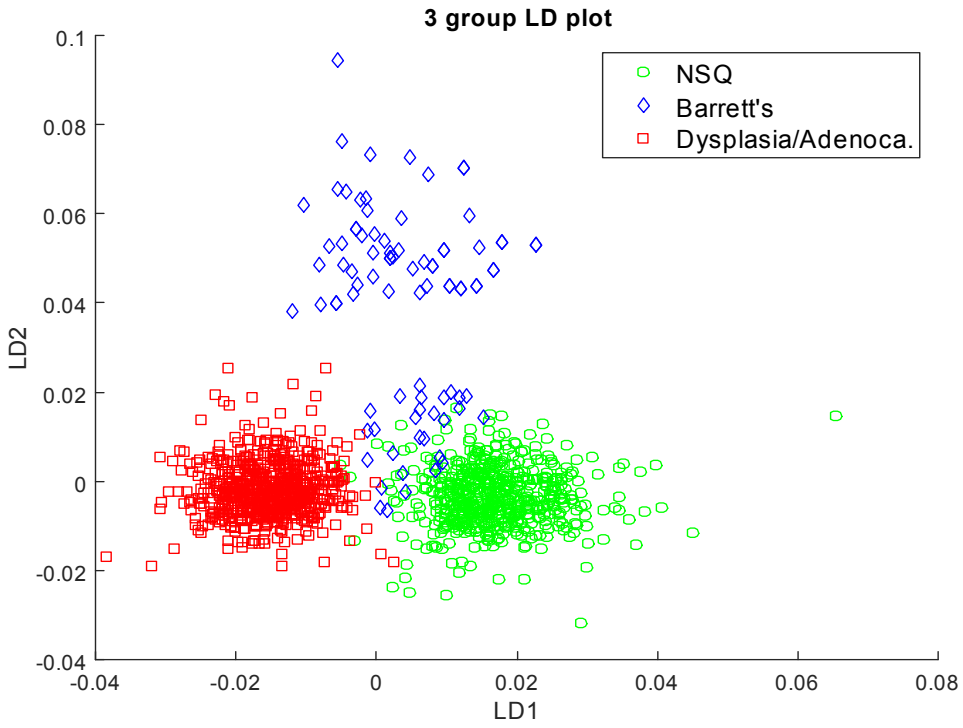


Figure 5-24 Scatter plot of all spectra, plotted by linear discriminant (LD) function. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A classification model was constructed using PCA-fed LDA, and tested using LOSOCV. The performance of the classification model is shown in Table 5-4.

The sensitivity of the model for individual spectra is reasonably good for normal squamous cells at 83.6%, but only moderate for the Barrett's and dysplasia/adenocarcinoma cells. The standard deviation was largest for the normal squamous cells, as would be expected since the cells chosen were

different for each iteration. The size of the variation in the sensitivity for the dysplasia/adenocarcinoma group reflects the overlap with the normal squamous cells and hence this varies depending on which cells are included. There was very little variation in the Barrett's result, with a small standard deviation.

Table 5-4 Performance of the training cells 3 group classification model

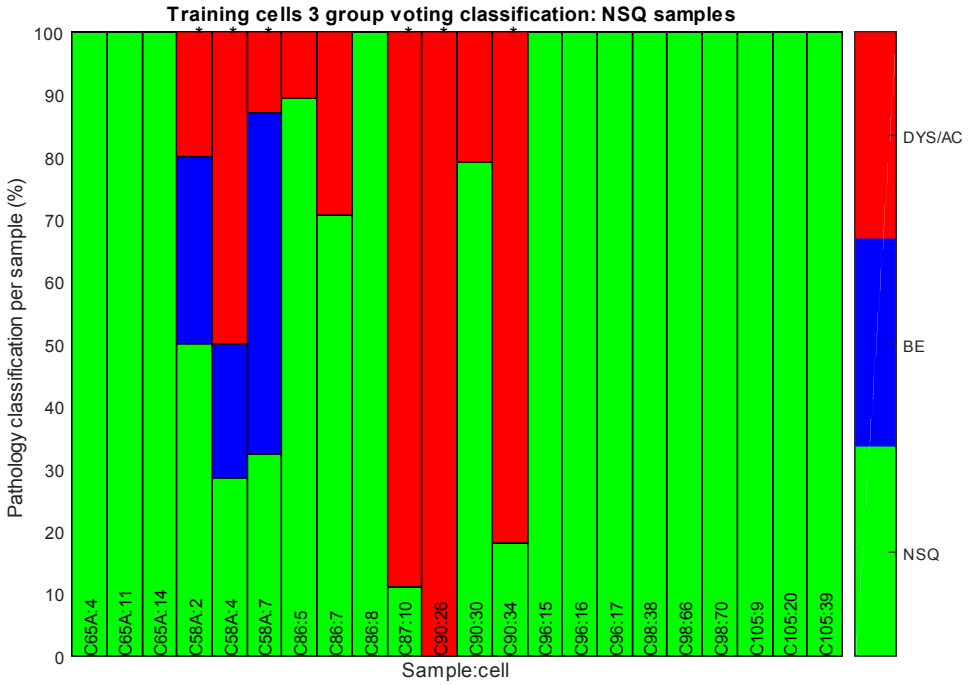
Mean values after 50 iterations		**NSQ Mean % (SD)**	**BE Mean % (SD)**	**DYS/AC Mean % (SD)**
Individual spectra	Sensitivity	83.6 (5.2)	62.8 (0.8)	69.5 (4.5)
	Specificity	70.8 (3.8)	97.5 (0.9)	87.8 (3.5)
30% threshold vote	Sensitivity	71.5 (5.3)	83.0 (1.7)	60.7 (7.1)
	Specificity	69.9 (4.9)	96.5 (2.0)	84.8 (3.5)
Mean spectra per cell	Sensitivity	61.9 (8.3)	78.5 (4.5)	43.0 (10.1)
	Specificity	59.9 (5.3)	94.9 (3.9)	77.9 (7.6)

SD standard deviation; NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/adenocarcinoma.

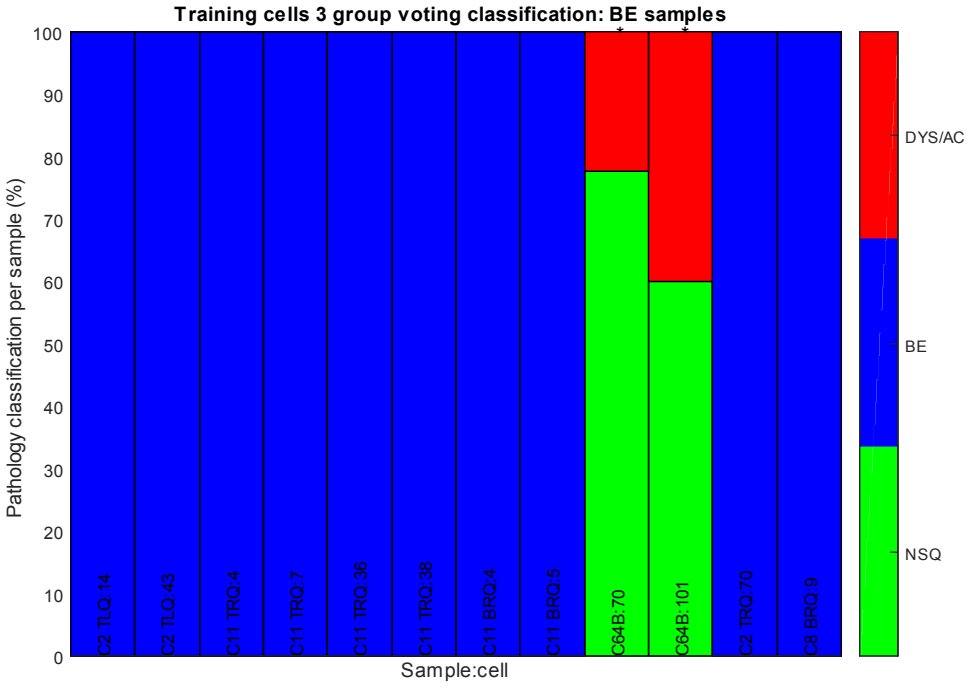
Classification of whole cells, rather than individual spectra, was attempted using a threshold voting system, as introduced for FTIR mapping samples in Chapter 4, and also using a training model developed using mean spectra per cell. The threshold voting system improved the result for Barrett's cells, but gave a poorer performance for the other two groups. This threshold voting tool was designed to give a high sensitivity for detecting pathology, but only achieved a detection

rate of 60.7% for dysplasia in this model. Analysis of the individual pixel classification sheds further light on this (see Figure 5-25).

A



B



C

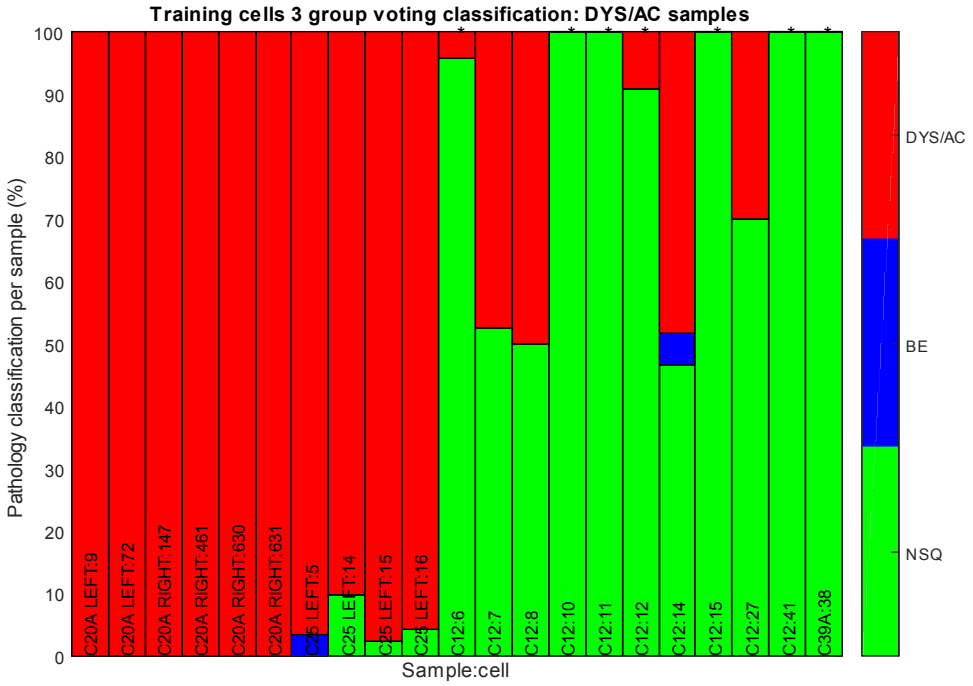


Figure 5-25 Classification of each cell using voting classification for 3 group model. The proportion of pixels classified as each pathology is shown for each cell within each pathology group in turn: A) normal squamous, B) Barrett's oesophagus, C) dysplasia/adenocarcinoma samples. Cells misclassified using the 30% threshold are marked with an asterisk at the top of the bar. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

For every cell from sample C12, over 50% of the pixels were misclassified, with the vast majority of these classified as normal squamous. Most of the cells from C12 were therefore classified as normal squamous, and since this sample contributed nearly 50% of the cells in the dysplasia/adenocarcinoma group, this had a strongly detrimental impact on the model. Possible reasons for this are considered in the discussion (section 5.14).

The Barrett's cells classified well in this model, but 2 cells from one sample were misclassified. In such a small sample size, this has a big impact on the classification performance.

The model built on mean spectra had an even worse performance than the majority voting, with sensitivity for dysplasia/adenocarcinoma 43.0%. Although not illustrated graphically, the model developed using mean spectra misclassified the same Barrett's and dysplasia/adenocarcinoma cells as the threshold voting based on individual spectra, but also misclassified the four C12 cells which were correct with threshold voting (and excluded one dysplastic cell entirely on EMSC correction). The normal squamous cells are randomly chosen and therefore change with each iteration.

Using mean spectra removes any weighting in the training model towards larger cells or clusters of cells, since each region contributes only one mean spectrum to the model. One potential cause of error is from pixels that include the edge of cells. This is due both to the increased Mie scattering that occurs at an edge, and because only part of the light for this pixel has passed through the cell, and some of the region contributing to the spectrum will be blank slide. These edge pixels may therefore have a lower quality signal than the pixels from the middle of a cell. For smaller cells the edge pixels will form a larger proportion of the total and the mean spectrum will be more affected by edge pixels.

One dysplastic sample in particular (C12) contributed heavily to the training dataset (10 out of 22 cells), but then performed very poorly in the classification model. On subsequent review of this sample, it was noted to be more heavily stained than many of the other samples. Since larger, darker nuclei are a feature of dysplastic change, heavy staining of the nuclei can potentially lead to

'over-calling' of cells as dysplastic. This is a possible explanation both for the high number of cells this sample contributed, and for its subsequent misclassification.

5.11.2. 2 group model

If cytology were to be used as a screening tool to identify Barrett's oesophagus (with patients then going on to have endoscopy) then discriminating between metaplasia and dysplastic change may not be required. Consequently, a two group analysis comparing normal squamous with a combined group of Barrett's, dysplasia and adenocarcinoma (BE/DYS/AC) was performed.

Once again, PCA-fed LDA was used to construct a model. The 3 most discriminatory PCs and LD functions are plotted in Figures 5-26 and 5-27. These figures show reasonably good grouping of the spectra within each group, and LD1 separates between the two groups with a few outlying spectra, and a small amount of overlap.

The model was then tested using LOSOCV: performance of the two group training model is shown in Table 5-5.

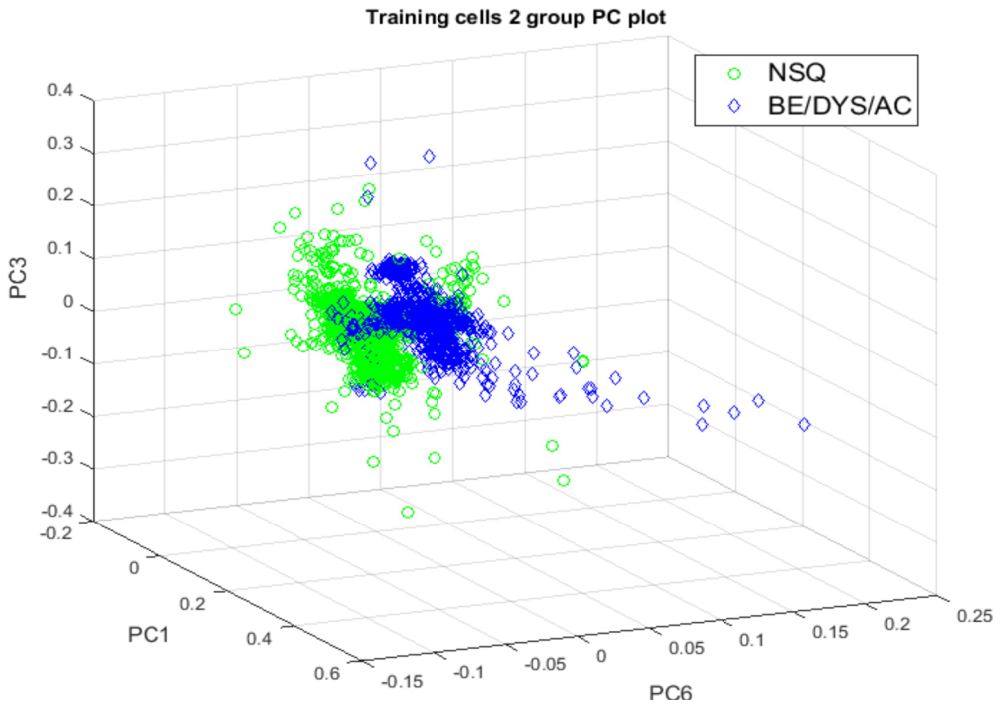


Figure 5-26 Scatter plot of all spectra, plotted by principal component (PC) score for 3 most discriminatory PCs: PC1, PC3 and PC6.

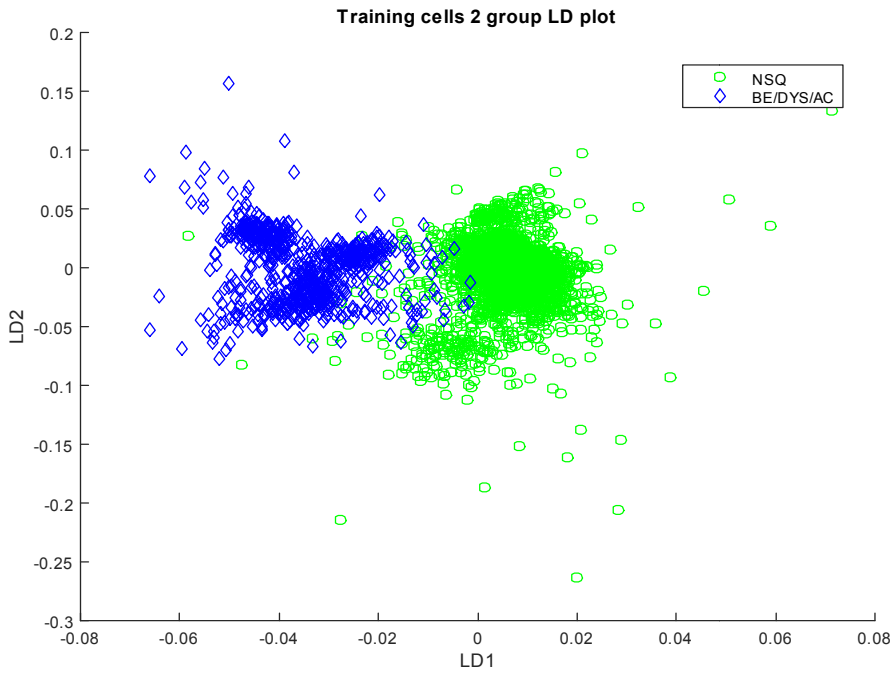


Figure 5-27 Scatter plot of all spectra, plotted by linear discriminant (LD) function. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

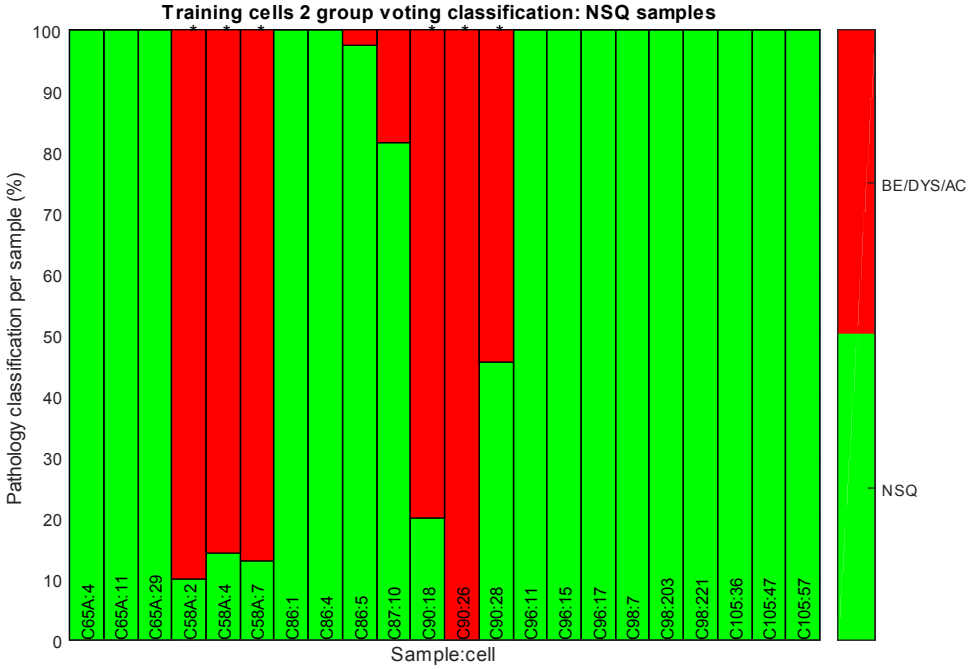
Table 5-5 Performance of the training cells in a two group training model

Mean values after 50 iterations		**NSQ Mean % (SD)**	**BE/DYS/AC Mean % (SD)**
Individual spectra	Sensitivity	**84.4 (5.8)**	**72.6 (2.6)**
	Specificity	**72.6 (2.6)**	**84.4 (5.8)**
30% threshold vote	Sensitivity	**71.6 (7.0)**	**66.0 (5.0)**
	Specificity	**66.0 (5.0)**	**71.6 (7.0)**
Mean spectra per cell	Sensitivity	**62.4 (11.5)**	**58.5 (5.8)**
	Specificity	**58.5 (5.8)**	**62.4 (11.5)**

Whilst the sensitivity for normal squamous cells is reasonably good at 84.4%, for the BE/DYS/AC group it is 72.6%. This figure is slightly better than for individual spectra using the 3 group model, but remains too low to be considered in clinical practice. Using the threshold vote and mean spectra, the detection rate is worse, with sensitivity of 66.0% and 58.5% respectively.

The threshold voting bar graphs in Figure 5-28 illustrate how certain samples are classified incorrectly. These poor results reflect the large contribution of sample C12 to the BE/DYS/AC group: as in the 3 group model, nearly every cell from this sample has been classified incorrectly. Since the dysplasia/adenocarcinoma group is larger than the Barrett's group, this single sample has a large overall effect.

A



B

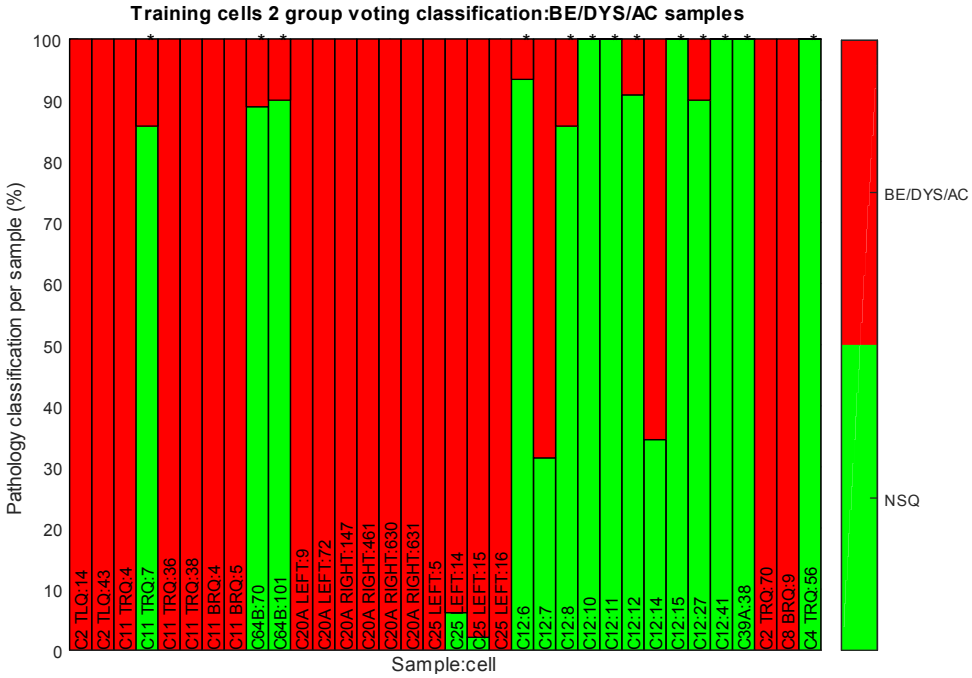


Figure 5-28 Voting classification of each cell for 2 group model. The proportion of pixels classified as each pathology is shown for each cell: A) NSQ B) BE/DYS/AC samples. Cells misclassified using the 30% threshold are marked with an asterisk at the top of the bar. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

5.11.3. Spectral analysis: spectral basis for classification

5.11.3.1. Spectral peak assignment

As discussed in the spectral analysis for the FTIR mapping chapter (section 4.10.1 above), assigning biomolecules to spectral peaks is tentative, as there are frequently several possible bond vibrations that can cause a peak at a given wavenumber, and multiple biomolecules that could explain a given bond vibration. Whilst interpreting the cell spectra, these problems are compounded by the much smaller numbers of spectra in each pathology group dataset than in the tissue work discussed earlier. Nonetheless, attempts to interpret the spectral differences are outlined below. The table of peak assignments listed in Section 4.10.1 (Table 4-9) is used again as a reference for spectral peak labelling.

As in the FTIR mapping section, inverse second derivative spectra were used to identify spectral peaks. The mean spectra for the 3 pathology groups included in the analysis are shown in Figure 5-29, along with the second derivative spectra in Figure 5-30.

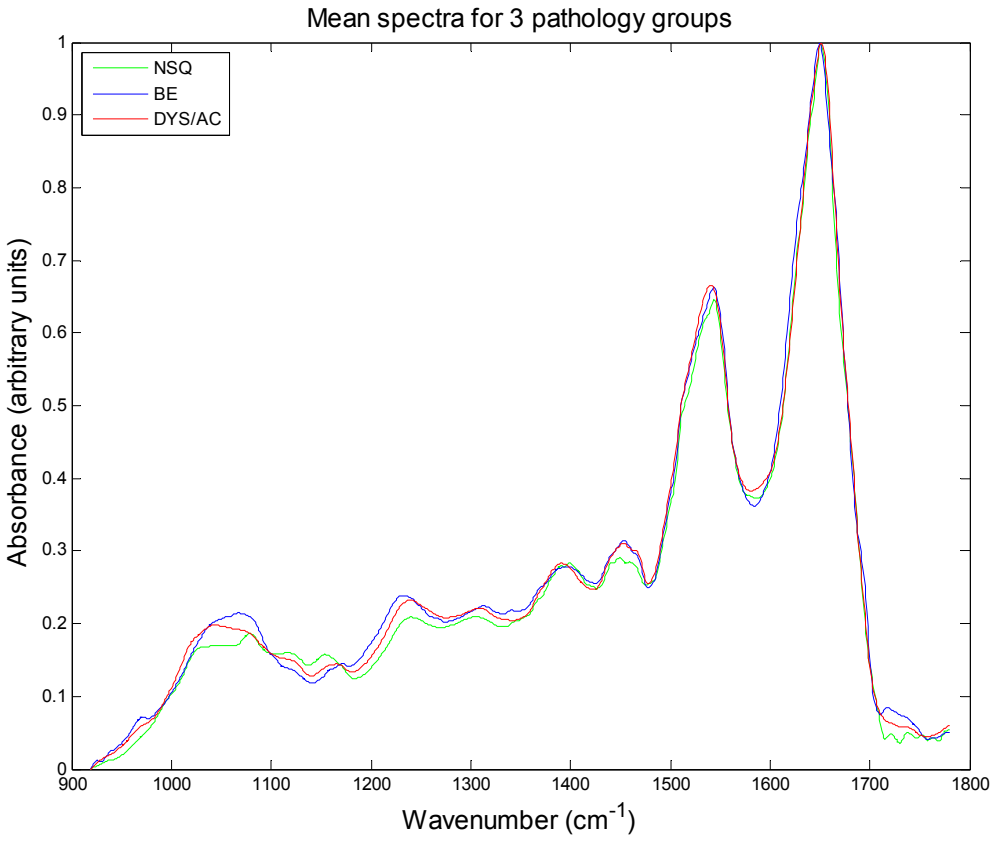
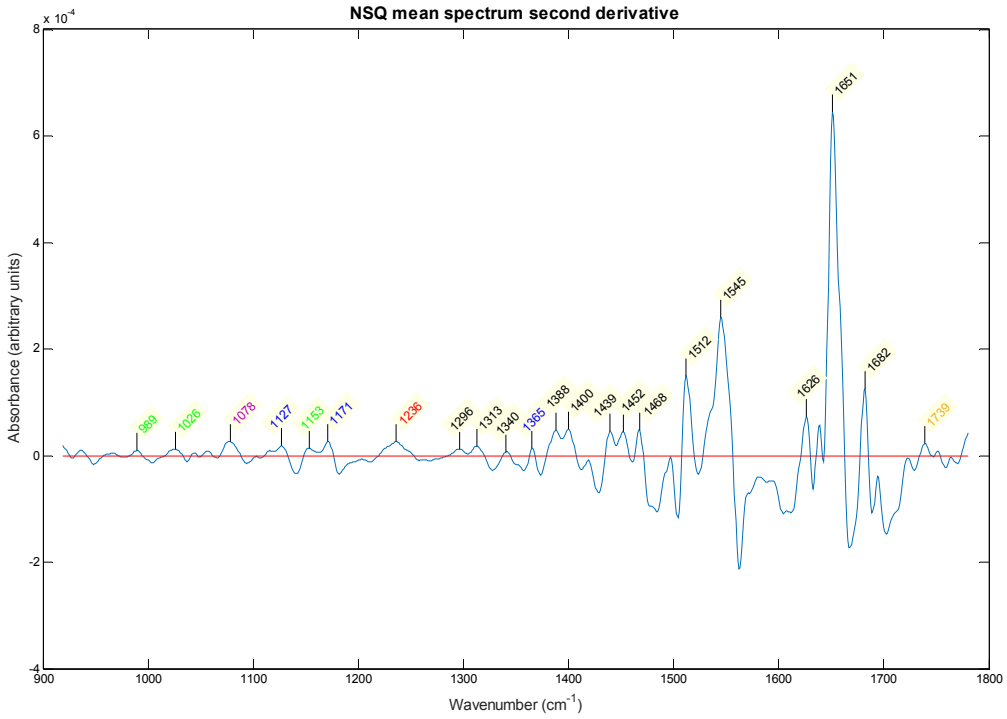
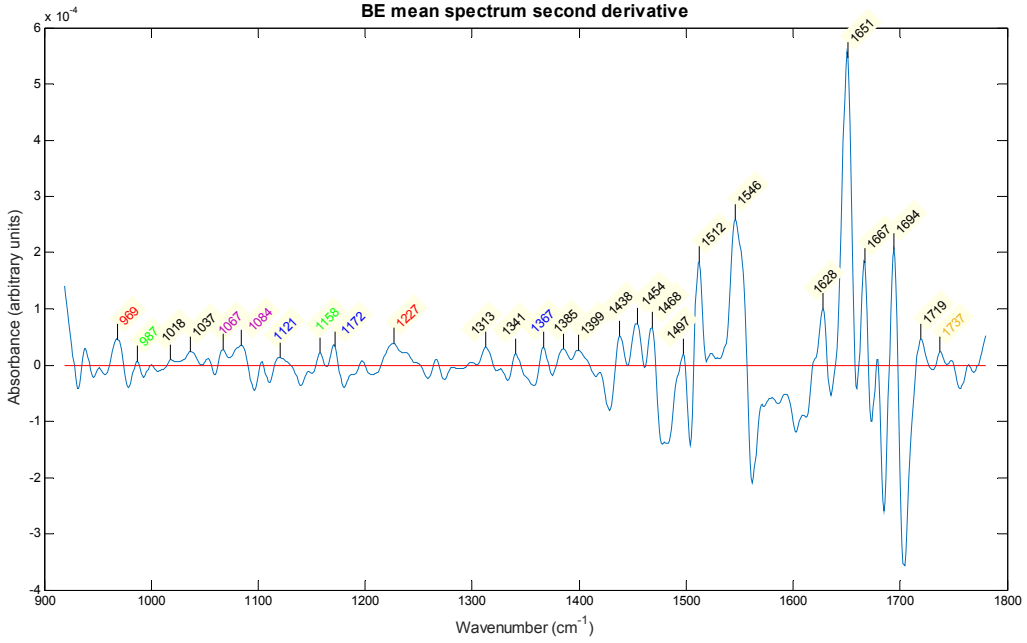


Figure 5-29 Mean spectra for all cellular regions included in the training dataset for the 3 pathology groups. NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/adenocarcinoma.

A


NSQ mean spectrum second derivative
x 10^-4
Absorbance (arbitrary units)
Wavenumber (cm^-1)
989
1026
1078
1127
1153
1171
1236
1296
1313
1340
1365
1388
1400
1439
1452
1468
1512
1545
1626
1651
1682
1739

B


BE mean spectrum second derivative
x 10^-4
Absorbance (arbitrary units)
Wavenumber (cm^-1)
969
987
1018
1037
1067
1084
1121
1158
1172
1227
1313
1341
1367
1385
1399
1438
1454
1468
1497
1512
1546
1628
1651
1667
1694
1719
1737

C

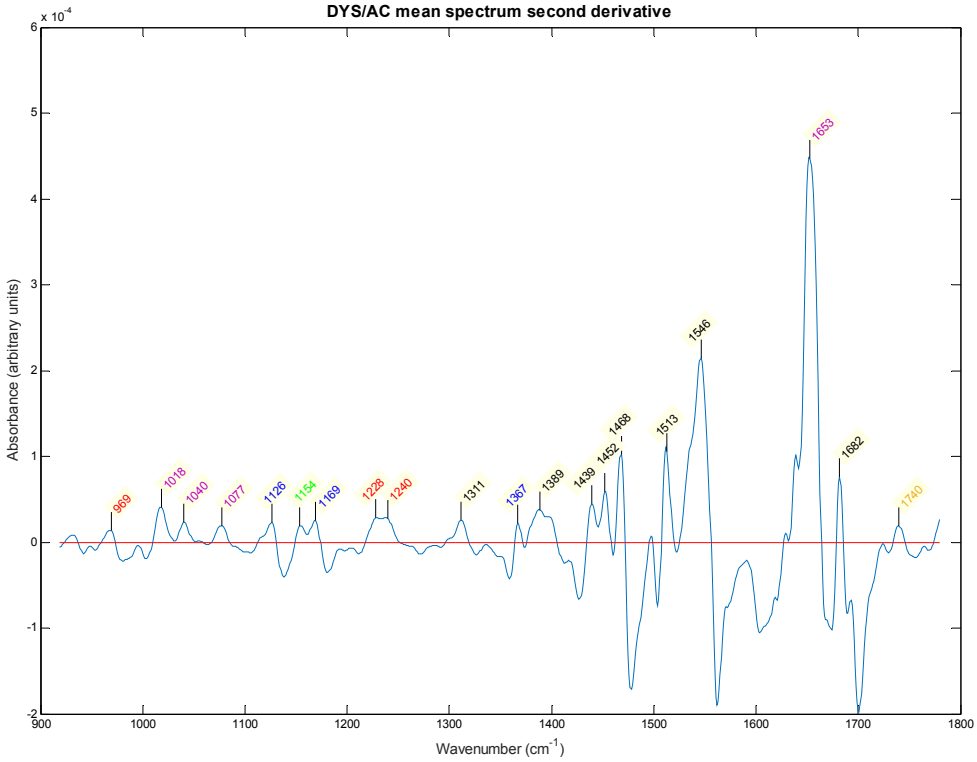


Figure 5-30 Second derivative of the mean spectra shown in Fig. 5-29, A) Normal squamous, B) Barrett's oesophagus, C) Dysplasia/adenocarcinoma, with the text colour indicating the likely biomolecule mainly responsible for that peak. NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/adenocarcinoma. Green = glycogen, blue = glycoprotein, red = DNA, Black = amino acid/protein, Purple = mixed contributions/multiple possibilities.

The major glycogen peaks identified in the FTIR mapping study described in section 4.10.1 occurred at 995, 1024 and 1153cm^{-1}; of these, the largest was the 1024cm^{-1} peak. The normal squamous inverse second derivative spectrum (Figure 5-30A) shows small peaks at 989, 1026 and 1153cm^{-1}, which likely correspond to glycogen as in the tissue study. The peak at 1078cm^{-1} may contain contributions from glycogen, glycoproteins and DNA (Quaroni and

Casson 2009; Wang et al. 2007). Small glycoprotein peaks occur at 1170 and 1365cm^{-1}: the peak at 1127cm^{-1} was not identified in the FTIR mapping study, but a peak at 1124cm^{-1} has previously been reported as representing glycoprotein in Barrett's samples by Quaroni and Casson (Quaroni and Casson 2009). The 1236cm^{-1} peak corresponds to the DNA peak seen in the FTIR mapping study at 1234-6cm^{-1}. Many peaks are likely to correspond to protein/amino acid peaks. The 1296 and 1313cm^{-1} peaks were not seen at this location in the mapping study, but a single peak at 1307cm^{-1} was attributed to phenylalanine, as was the 1338cm^{-1} peak (likely to be the 1340cm^{-1} peak seen here). Similarly, there was a 1394cm^{-1} protein peak seen in the mapping study, whereas there is now a doublet at 1388 and 1400cm^{-1} in the cell spectra. Further protein peaks around 1450cm^{-1}, the amide I and amide II regions, are again similar to the mapping study. The small peak at 1739cm^{-1} may correspond to phospholipid, where the peak was seen at 1741-7cm^{-1} in the mapping study.

The Barrett's oesophagus inverse second derivative spectrum is shown in Figure 5-30B. There are many similarities to the normal squamous spectrum and a number of differences. There is an additional peak at 969cm^{-1} which probably represents DNA, with a higher and broader peak at 1227cm^{-1} than the normal squamous DNA peak at 1234cm^{-1}. The peaks at 987 and 1158 may represent glycogen, though they are very small, but the peak at 1024-6cm^{-1} is missing. Instead there are peaks at 1018 and 1037cm^{-1} that are not seen in the normal squamous group or the tissue mapping study. Whilst this may simply reflect a different location for glycogen peaks, there is an appreciable difference in the raw spectra in Figure 5-29, between wavenumbers 1000-1090. This shows a broader, higher intensity peak in the Barrett's and

dysplasia/adenocarcinoma group than the normal squamous cells. Since a higher proportion of the Barrett's/dysplasia/adenocarcinoma samples were preserved in formalin, and the formalin spectrum shows a strong peak in this region, the possibility of a contribution from the formalin solution has also been considered. This is discussed further below, in section 5.11.4.

Whilst there is a single peak at 1078cm^{-1} in normal squamous cells, there are two peaks at 1067 and 1084cm^{-1} in the Barrett's cells. This may follow from the different contributions of glycoprotein, glycogen, and DNA in this region. There are similar glycoprotein peaks at 1121, 1172 and 1367cm^{-1}, though with greater intensities in the Barrett's cells. The protein and amino acid peaks in the amide I, amide II and amide III regions are all at very similar wavenumbers to those in the normal squamous cells, with the exception of the region 1667-1694cm^{-1}. The principal amide I peak at 1651cm^{-1} has greater intensity in the normal squamous group, which correlates with the finding in the tissue mapping study.

The dysplasia/adenocarcinoma inverse second derivative spectrum has many similarities to the other groups. It has a DNA peak at 969cm^{-1} as seen with the Barrett's cells, and the peak around 1234cm^{-1} is broader, stretching between 1228-1240cm^{-1}. Peaks at 1018 and 1040cm^{-1} are seen as in the Barrett's cells, with no glycogen peak at 1026cm^{-1} as in the normal squamous group and in the tissue mapping study. The only clear glycogen peak is at 1154cm^{-1}. Glycoprotein peaks occur at 1126, 1169 and 1367cm^{-1}, similar to the other 2 groups. The protein peaks in the amide I, amide II and amide III regions are similar to the other groups, though the amide I region is perhaps more similar to the normal squamous rather than the Barrett's cells.

In summary, there are many similarities in the spectral peaks seen in each of the three cell groups to the peaks identified in the tissue mapping section earlier. The amide I (1651-3cm^{-1}) and amide II (1546cm^{-1}) peaks are in almost exactly the same positions as the tissue mapping (1651-3 and 1547cm^{-1} respectively). A DNA peak was seen at 1234-6cm^{-1} in the tissue mapping, and there is a peak in this region for each of the cell groups, though with some variation. There were some minor differences in the peaks attributed to glycoprotein – with peaks 1121-7cm^{-1} (1113-7cm^{-1} in the tissue work), 1169-72cm^{-1} (1171-2cm^{-1} in tissue), and 1365-7cm^{-1} (1373-7cm^{-1} in tissue). The peak seen at 1080cm^{-1} in the tissue study was attributed to mixed contributions of glycogen, glycoprotein and DNA: the normal squamous and dysplasia/adenocarcinoma groups showed a peak at 1077cm^{-1}, but the Barrett's group had a doublet in this region at 1067 and 1084cm^{-1}. The glycogen peaks also show a difference: in the tissue work, glycogen peaks were seen at 995, 1022-6 and 1153-7cm^{-1}. In the cell spectra, peaks were seen at 987-9, 1026 and 1153-8cm^{-1}. However, the 1026cm^{-1} peak was seen only in the normal squamous cells, and not in the Barrett's or dysplasia/adenocarcinoma groups, both of which showed a doublet in this region of 1018 and 1037-40cm^{-1}.

The phospholipid peak seen around 1740-5cm^{-1} in the tissue mapping study is not seen here. It is possible that signal from cell membranes may be removed during pre-processing. The pixels at the edge of cells can contribute artefact due to scattering effects: if this were consistently recognised by the pre-processing steps they might therefore be removed.

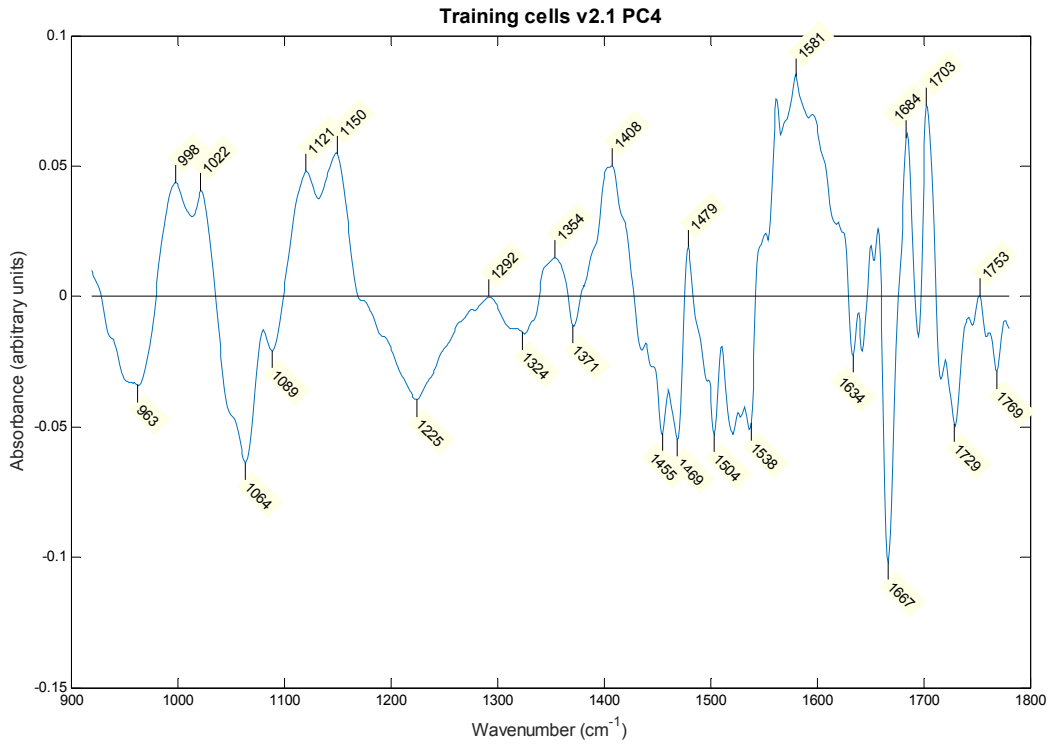
The many similarities between the cell spectra and the earlier tissue study support the findings of both studies, and imply that these spectral peak positions will be consistent findings in both oesophageal cells and tissue.

5.11.3.2. Principal component loadings analysis

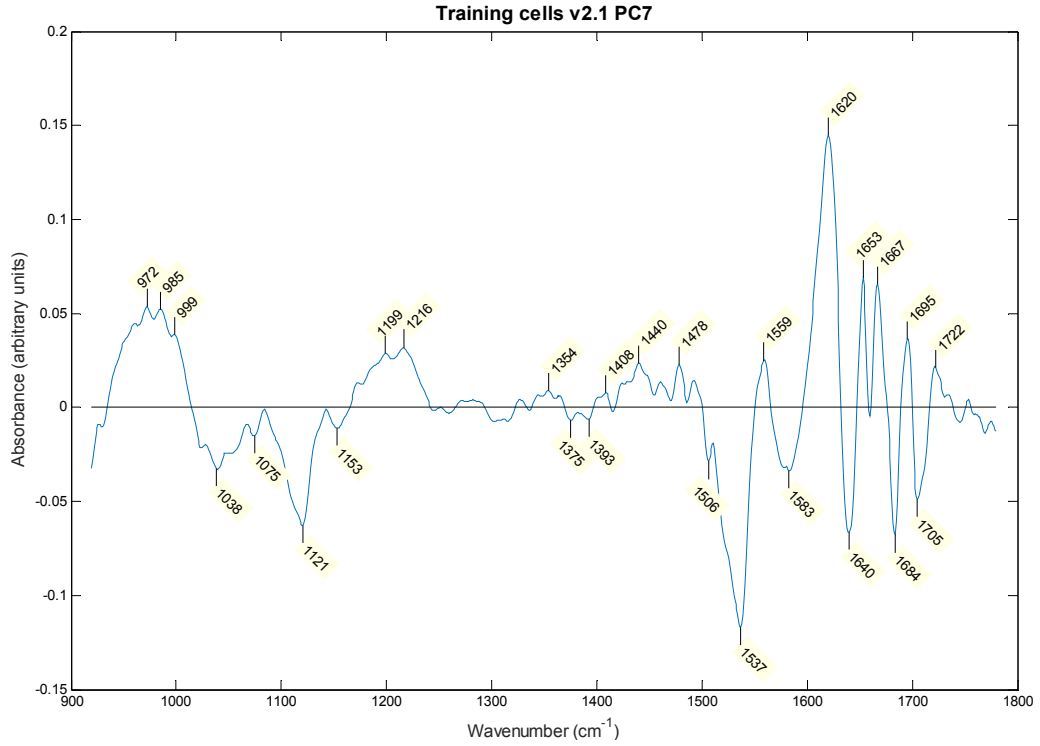
In order to analyse which spectral differences have contributed to the classification model, the most discriminatory principal component (PC) loadings were analysed. The 5 most discriminatory PCs are shown in Figure 5-31.

Many of the key discriminatory peaks seen in the tissue mapping study, and identified in the inverse second derivative spectra of the cells, do not feature prominently in the PC loadings. The glycogen peak at 1022-6cm^{-1}, which occurred only in the normal squamous cells, appears only in PC4. The glycogen peak at around 1150cm^{-1} appears in PC4 (1150cm^{-1}), PC7 (1153cm^{-1}), and may be part of a broader peak in PCs 3 (1140-1203cm^{-1}) and 11 (1145-89cm^{-1}). However, the other glycogen peak seen around 987-9cm^{-1} does not appear consistently at this position in the PC loadings, and in the absence of the other glycogen peaks it is difficult to draw conclusions from the presence of the 1150cm^{-1} peak alone. Where it occurs with the 1022cm^{-1} peak in PC4 (both positive), this suggests that high glycogen content is selected for in the positive direction. Thus PC4 could potentially be separating out the normal squamous cells. This is supported by the negative peaks, which correspond to the doublet seen at 1064 and 1089cm^{-1} in the Barrett's cells, along with the DNA peak at 1225cm^{-1} (the position for the Barrett's cells) and the peak at 1371cm^{-1} which may represent glycoprotein. Taken together, this implies that PC4 may be separating the normal squamous (positive) from Barrett's (negative) cells.

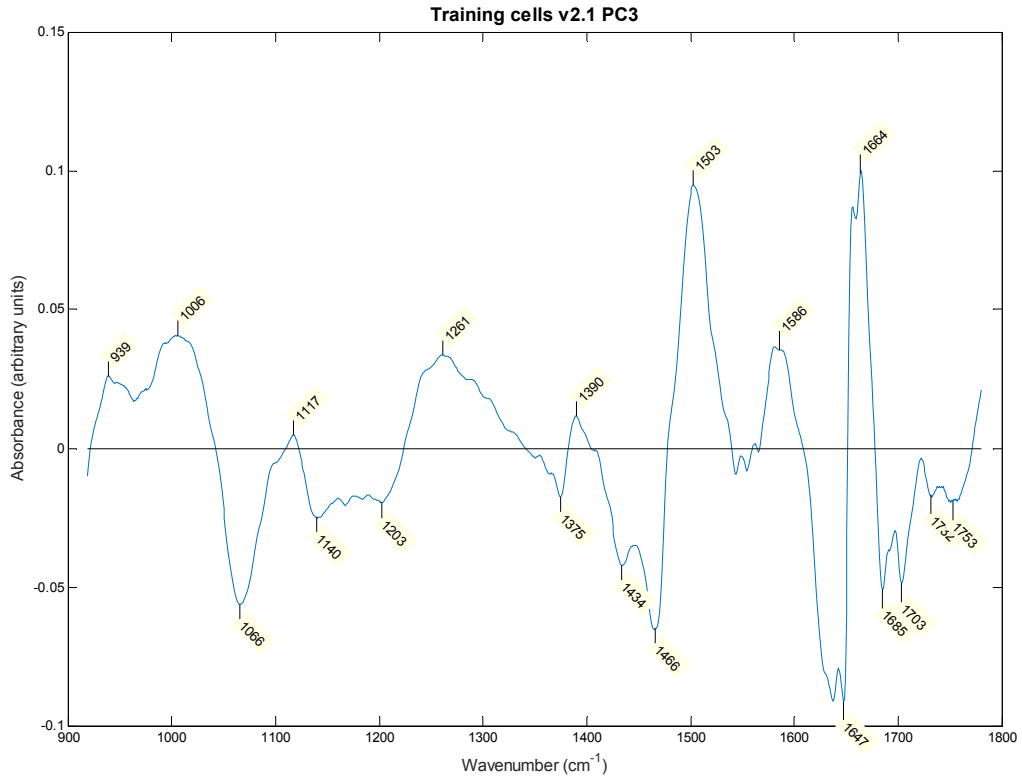
A


Training cells v2.1 PC4
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.1
0.05
0
-0.05
-0.1
-0.15
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
963
988
1022
1064
1089
1121
1150
1225
1292
1324
1354
1371
1408
1455
1469
1479
1504
1538
1581
1634
1667
1684
1703
1729
1753
1769

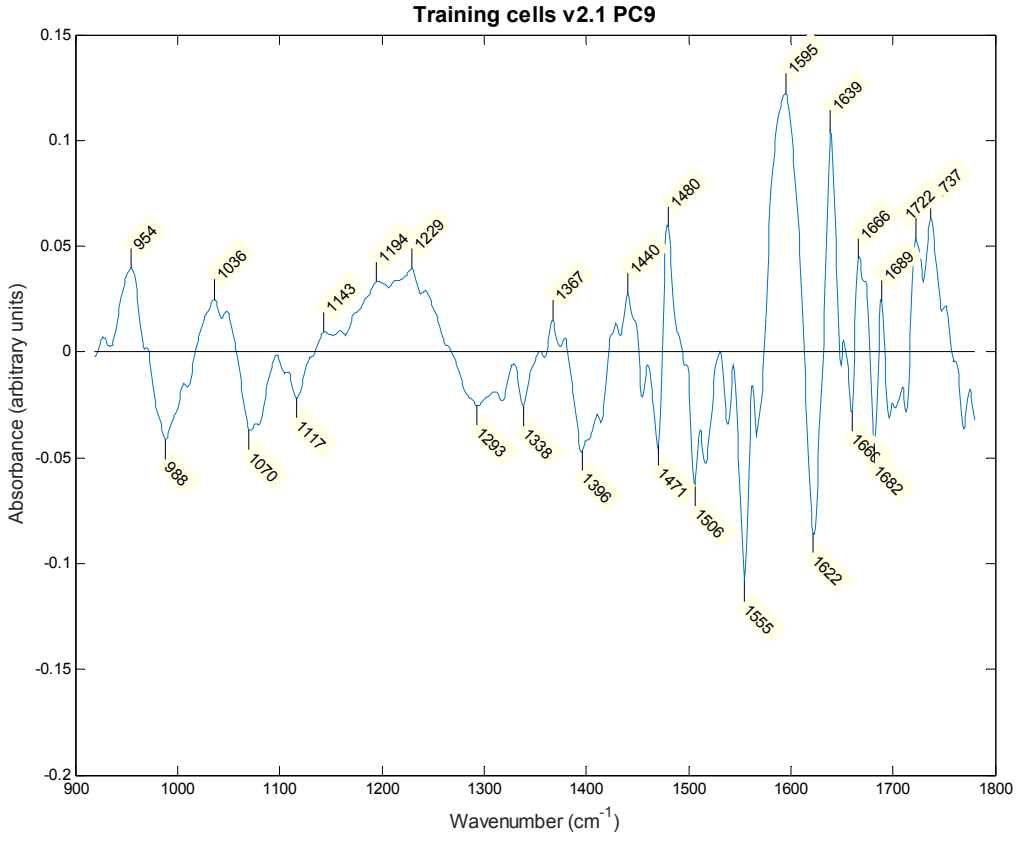
B


Training cells v2.1 PC7
Absorbance (arbitrary units)
Wavenumber (cm-1)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
972
985
999
1038
1075
1121
1153
1199
1216
1354
1375
1393
1408
1440
1478
1506
1537
1559
1583
1620
1640
1653
1667
1684
1695
1705
1722

C


Training cells v2.1 PC3
Absorbance (arbitrary units)
Wavenumber (cm-1)
939
1006
1066
1117
1140
1203
1261
1375
1390
1434
1466
1503
1586
1647
1664
1685
1703
1732
1753

D


Training cells v2.1 PC9
Absorbance (arbitrary units)
Wavenumber (cm-1)
954
988
1036
1070
1117
1143
1194
1229
1293
1338
1367
1396
1440
1471
1480
1506
1555
1595
1622
1639
1666
1682
1689
1722
737

E

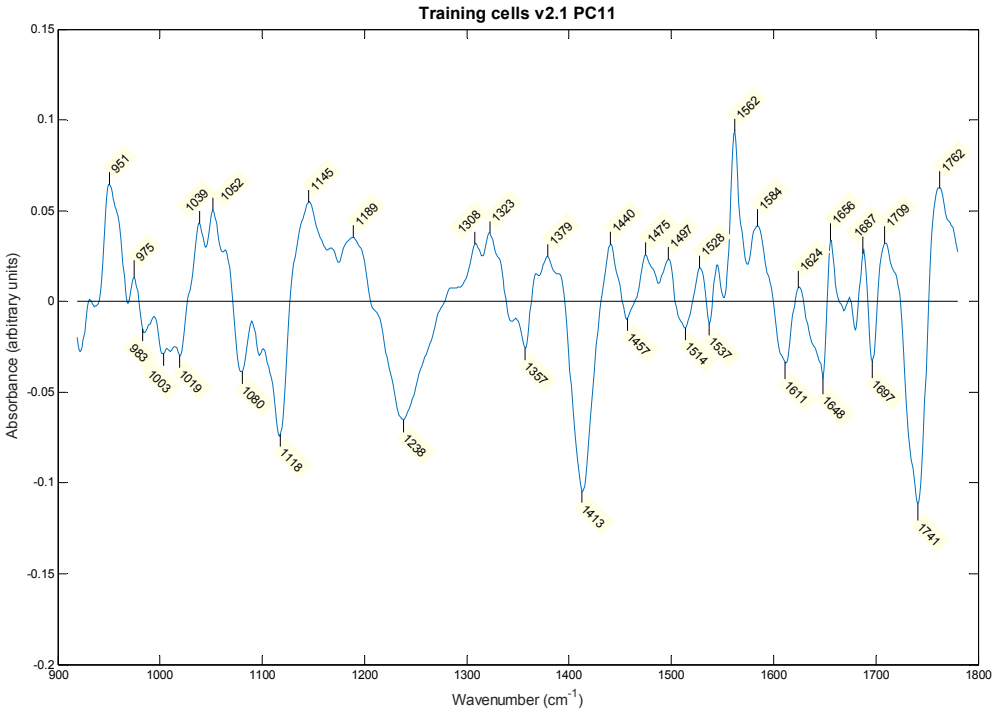


Figure 5-31 Most discriminatory PC loadings in descending order of discrimination. A) PC4, B) PC7, C) PC3, D) PC9, E) PC11.

The discriminatory role of the other PCs is less clear. Whilst certain peaks feature frequently (e.g. 1367-75cm^{-1} may be glycoprotein), in the absence of further peaks representing the same biomolecule to support this assertion, conclusions are limited. The major DNA peak which was a strong discriminator in the tissue work (1234-6cm^{-1}) does not appear frequently in these PC loadings. There is a large peak at 1238cm^{-1} in PC11 which may represent DNA (this peak is seen at 1234cm^{-1} in the normal squamous cells inverse second derivative spectra, and a broader peak is seen at this location for the dysplasia/adenocarcinoma group), and the 1225cm^{-1} peak in PC4 may represent DNA in the Barrett's cells as mentioned above.

The protein peaks in the range 1400-1700cm^{-1} generally do not occur at consistent positions and are therefore difficult to interpret.

The peak at 1036-9cm^{-1} occurs in 3 of the PC loadings (PCs 7, 9 and 11). The biomolecular origin for this peak is not entirely clear, as discussed above, and if artefact, this may represent a significant source of variance in the data, and is a major potential source of error.

Whilst there are a number of spectral peaks seen in the PC loadings that correspond to peaks identified in the inverse second derivative spectra, many of these peaks do not feature strongly, and there are other peaks present that do not correspond well with the key discriminatory peaks seen in the tissue work. This suggests that the spectral basis of classification for the cells may not have included certain key biochemical differences between the pathology groups, which could contribute to a poor classification performance. Simultaneously, the PC loadings may include artefact/non-pathological differences between the groups, which could give a falsely elevated performance amongst the training cells, but would likely reduce the performance if applied to new test data. These factors may be a result of the small sample size and hence sample differences may be more significant source of variation than any true biochemical differences resulting from pathology.

5.11.4. Revisiting the spectral contribution of CytoLyt™ vs. formalin

The spectral analysis revealed that there were some differences between the groups in the wavenumber region 1000-1100cm^{-1}, that did not correspond with the findings in the earlier tissue mapping study. One particular unexpected

finding was the broad peak seen in the Barrett's and dysplasia/adenocarcinoma group that was not present in the normal squamous cells (see Figure 5-32).

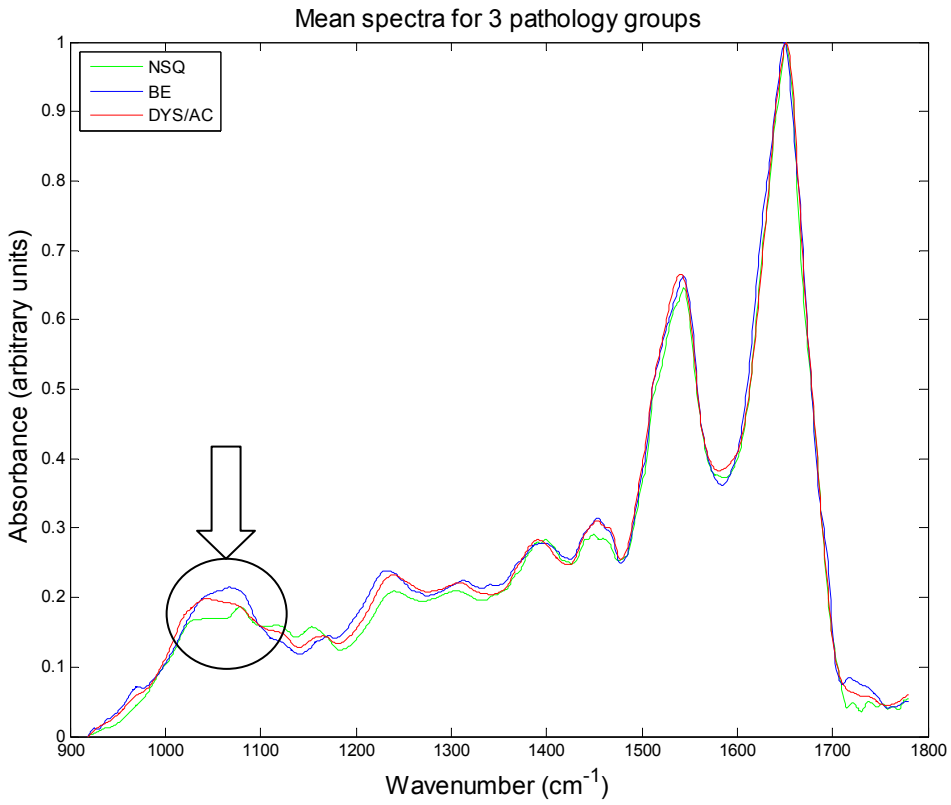


Figure 5-32 Region of unexpected difference between the pathology groups

Since this was not seen previously, we considered possible causes for this difference other than cell pathology. One other difference between the groups was that many more of the NSQ samples were preserved in CytoLyt™ (rather than formalin) compared to the Barrett's and dysplasia/adenocarcinoma samples (1 out of the 4 Barrett's samples and none of the 5 dysplasia/adenocarcinoma patients' samples were stored in CytoLyt™, versus 4 out of 8 normal squamous samples). Although principal component analysis

had failed to find a consistent difference between cells prepared using the two different fixatives, we nonetheless revisited the possibility that this could be the cause of a spectral difference.

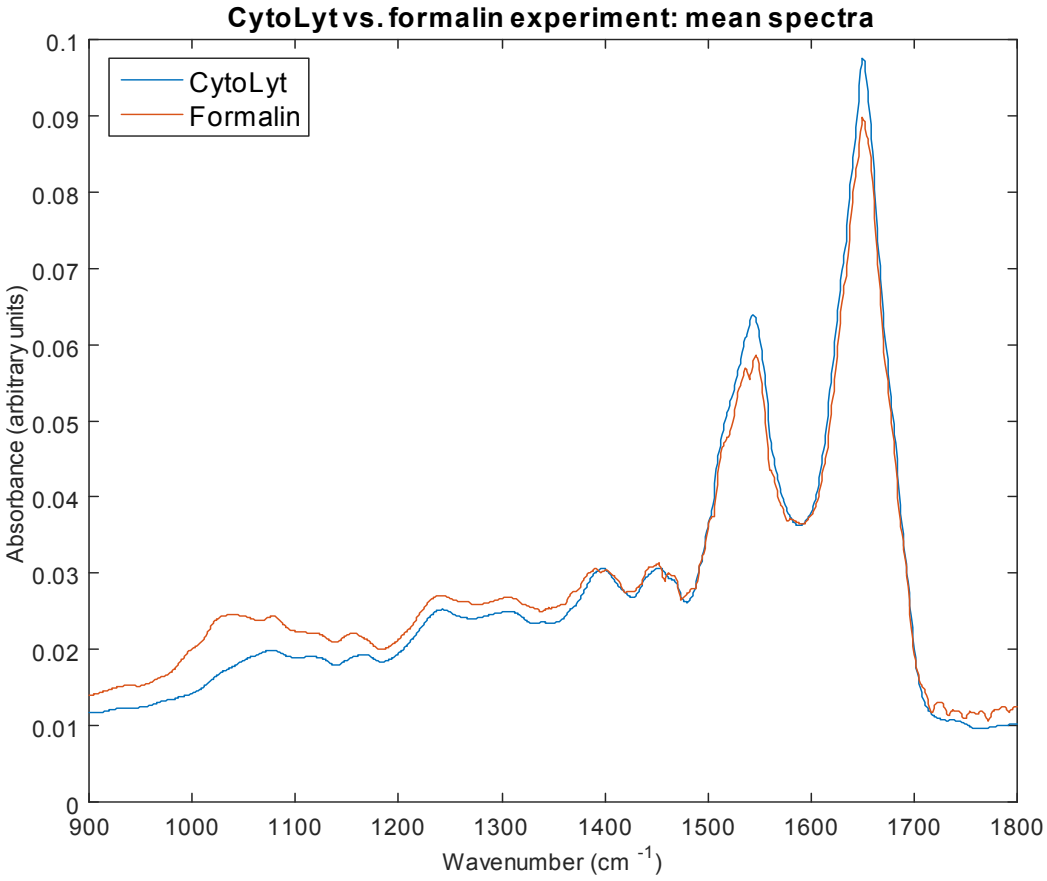


Figure 5-33 Mean spectra from earlier experiment comparing Cytolyt™ and formalin

The initial experiment comparing the fixative solutions compared 10 samples, 5 stored in formalin, 5 in CytoLyt™. The mean spectra of the two groups showed a difference in this same region between 1000-1100cm^{-1} (see Figure 5-33), though this was initially attributed to baseline differences seen from 900-1200cm^{-1}. Reassessing this with the experience gained from the training cells analysis, it appeared that the difference seen in the training cell groups

collected in formalin (Barrett's/dysplasia/adenocarcinoma), i.e. a broad peak starting around 1000cm^{-1} and extending to 1080cm^{-1}, was also present in the formalin group in the initial experiment.

To explore this further, we went on to examine the mean spectra for cells in the training dataset collected in formalin versus CytoLyt™. Since half of the normal squamous cell samples in the training dataset were collected in each fixative, we divided the normal squamous cells into 'formalin' and 'CytoLyt™' groups for comparison. The mean spectra are plotted in Figure 5-34.

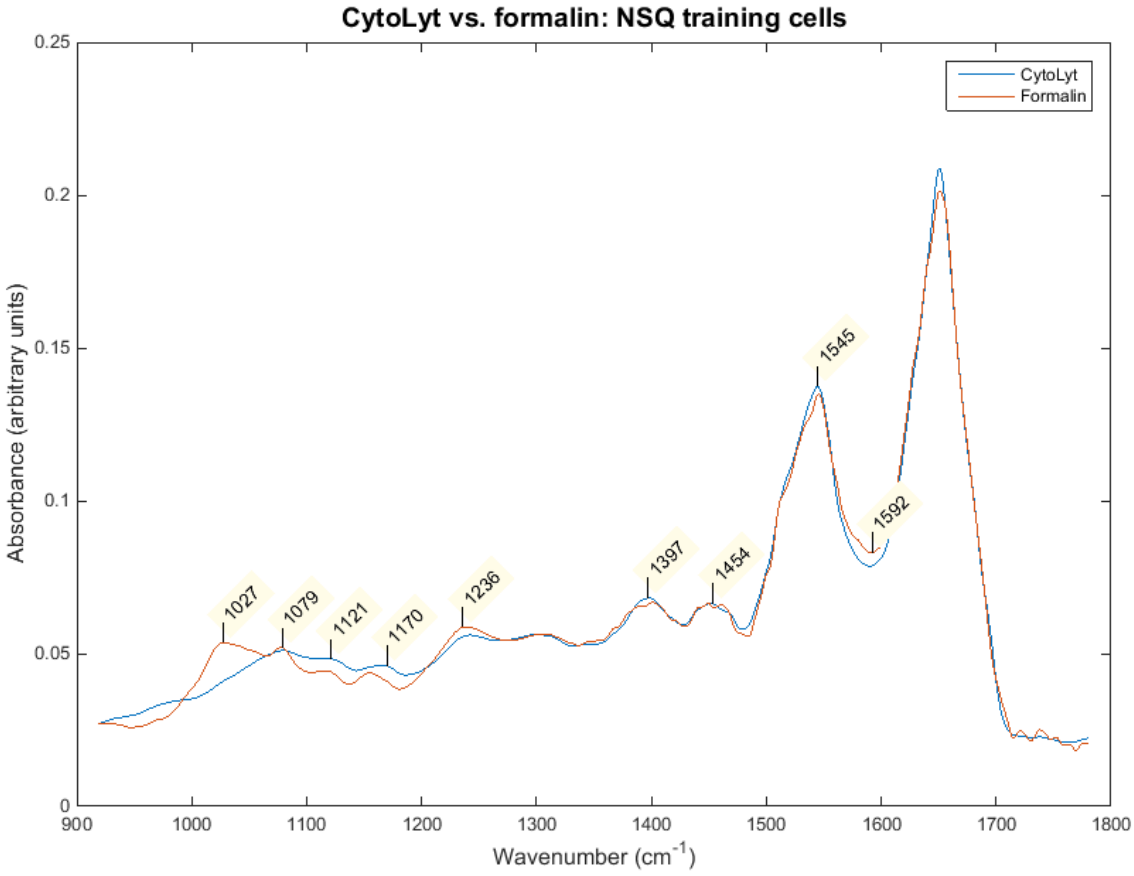


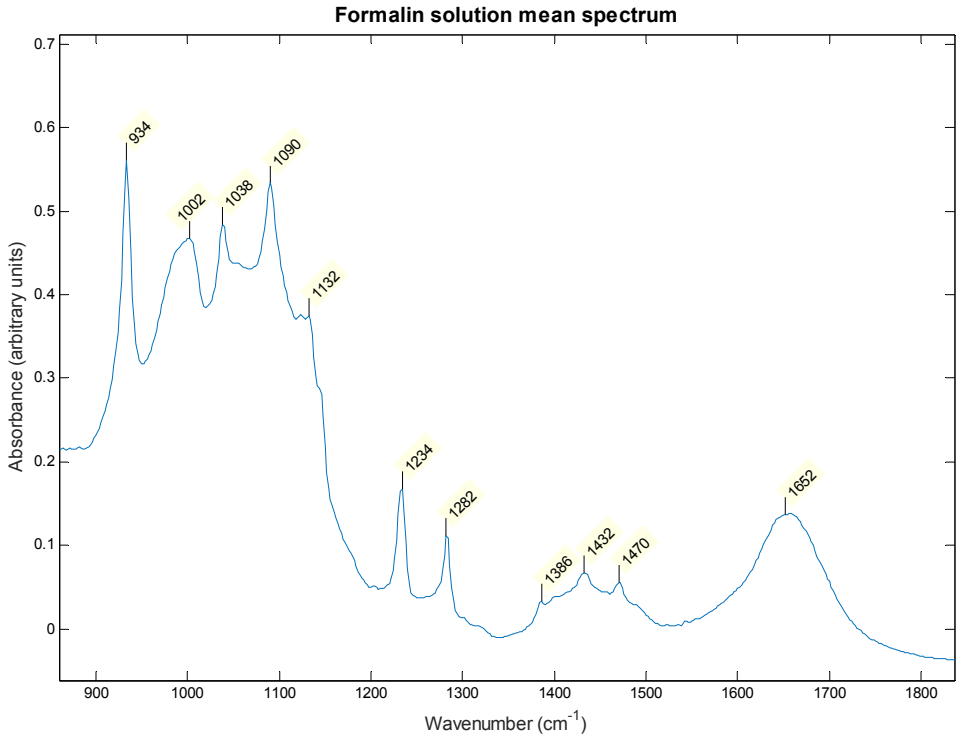
Figure 5-34 Comparison of mean spectra for all normal squamous cells in the training dataset, based on fixative solution used.

Once again, there is a difference in the same region, with the cells stored in formalin showing a broad peak around 1000-1080cm^{-1} that is not present in the cells stored in CytoLyt™.

Given this consistent difference in the spectra in this region, the raw fixative solutions were measured to see if such a difference due to the solutions was plausible. A drop of solution was placed on a clean calcium fluoride slide and spectral measurements taken, using the same instrumentation and settings as the cell measurements. The spectra from the raw solutions are shown in Figure 5-35.

Looking at the spectra from the raw fixative solutions, the formalin solution has its strongest signal across the wavenumbers 934-1090cm^{-1}, with various peaks in this range. The CytoLyt™ does not have strong signal in this region, but shows large peaks at 1424 and 1572cm^{-1}. It is difficult to predict how this would translate into the measured spectrum of a cell containing a small amount of this solution, since there would be many other compounds present that could influence the spectrum. The region of strongest signal in the formalin solution does not correspond perfectly with that seen in the cells stored in formalin, and the strongest features in the CytoLyt™ solution are not clearly seen in the cells stored in CytoLyt™. However, given the consistent finding that cells stored in formalin have a stronger signal in the range 1000-1080cm-1, and this wavenumber region occurs in the region of strongest signal for formalin solution, it certainly seems possible that this is a potential cause for the difference observed.

A



B

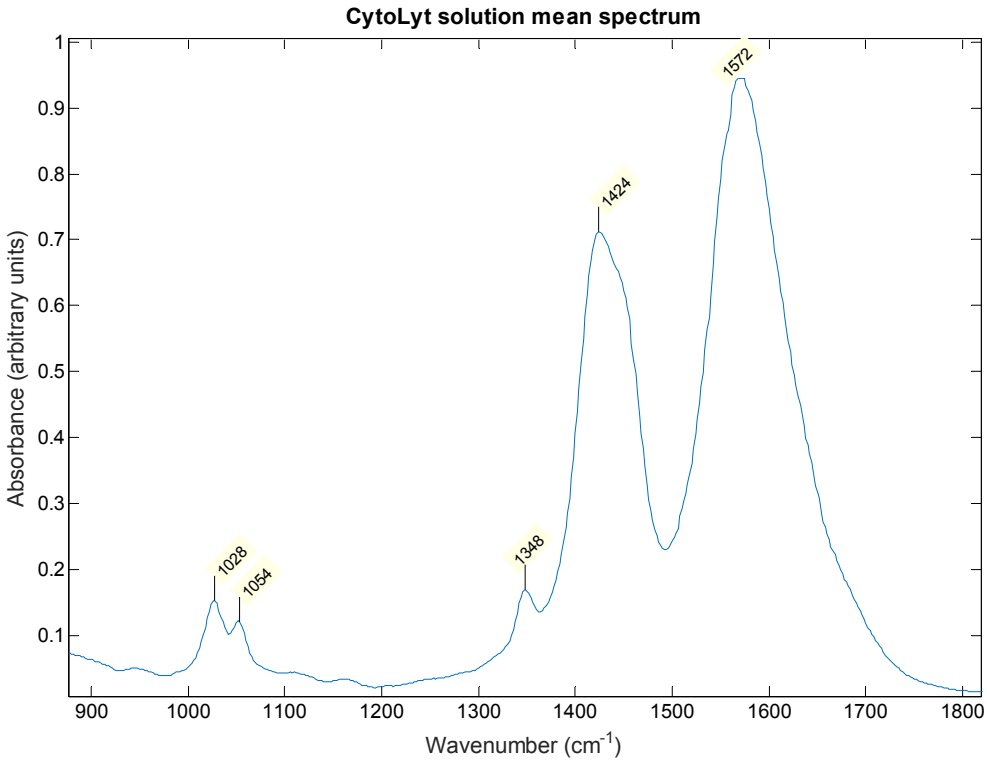


Figure 5-35 Mean spectra of raw solutions A) Formalin and B) CytoLyt™.

A training model developed using cells only stored in one type of fixative (preferably CytoLyt™) would therefore be desirable. Given the moderate classification performance of our training dataset, with the major limitation being sample size in each group, subgroup analysis of our dataset to include only cells stored in one type of fixative is not appropriate, as firm conclusions cannot be drawn.

5.12. Whole sample classification

The aim of this study was to develop a classification model using cytopathologically identified cells, which could then be applied to whole cell samples. As discussed above, the performance of the training cell model was only moderately good, and spectral analysis suggested that certain features were not included in the model, and differences due to artefact may have been included e.g. difference due to formalin preservative. In addition, the variability of the plating had resulted in large numbers of samples having no identifiable cells of the 'correct' pathology. Given these limitations, conclusions from the application of the training model to the whole sample test set may be limited. Nonetheless we proceeded to complete the analysis, not least because the process of developing the software algorithms is an important next step in the process.

The number of samples and spectra included in the test dataset (after the same pre-processing steps as those applied to the training dataset) are shown in Table 5-6. Cells that were included in the training dataset were excluded from the test set, to prevent overestimation of performance.

Table 5-6 Total number of samples included in the test dataset

	No. of patients	No. of FTIR maps	No. of cell regions	No. of spectra for analysis
Normal Squamous	18	19	722	27,662
Barrett's oesophagus	21	33	1,891	48,322
Dysplasia/ adenocarcinoma	25	42	2,620	54,418
Stomach	21	21	969	33,032
TOTAL	66*	115	6,202	163,434

*Some patients contributed to both stomach and one other pathology group

The stomach samples were included in the test dataset, even though they did not form part of the training set, simply as an exploratory experiment to see how they would classify amongst the other three groups. Without a training group, they cannot be classified 'correctly' as there is no 'stomach' group, but it is of interest to see whether they are spectrally more similar to the other non-pathological group (normal squamous) or the other glandular groups (Barrett's/dysplasia/adenocarcinoma).

The Barrett's and dysplasia/adenocarcinoma groups have greater numbers of samples: because of the anticipated difficulty in identifying representative cells in these pathology groups, certain samples had more than one region measured to increase the number of cells that could contribute to the training set. These different regions on the same slide will be considered as separate samples for the purposes of this test classification.

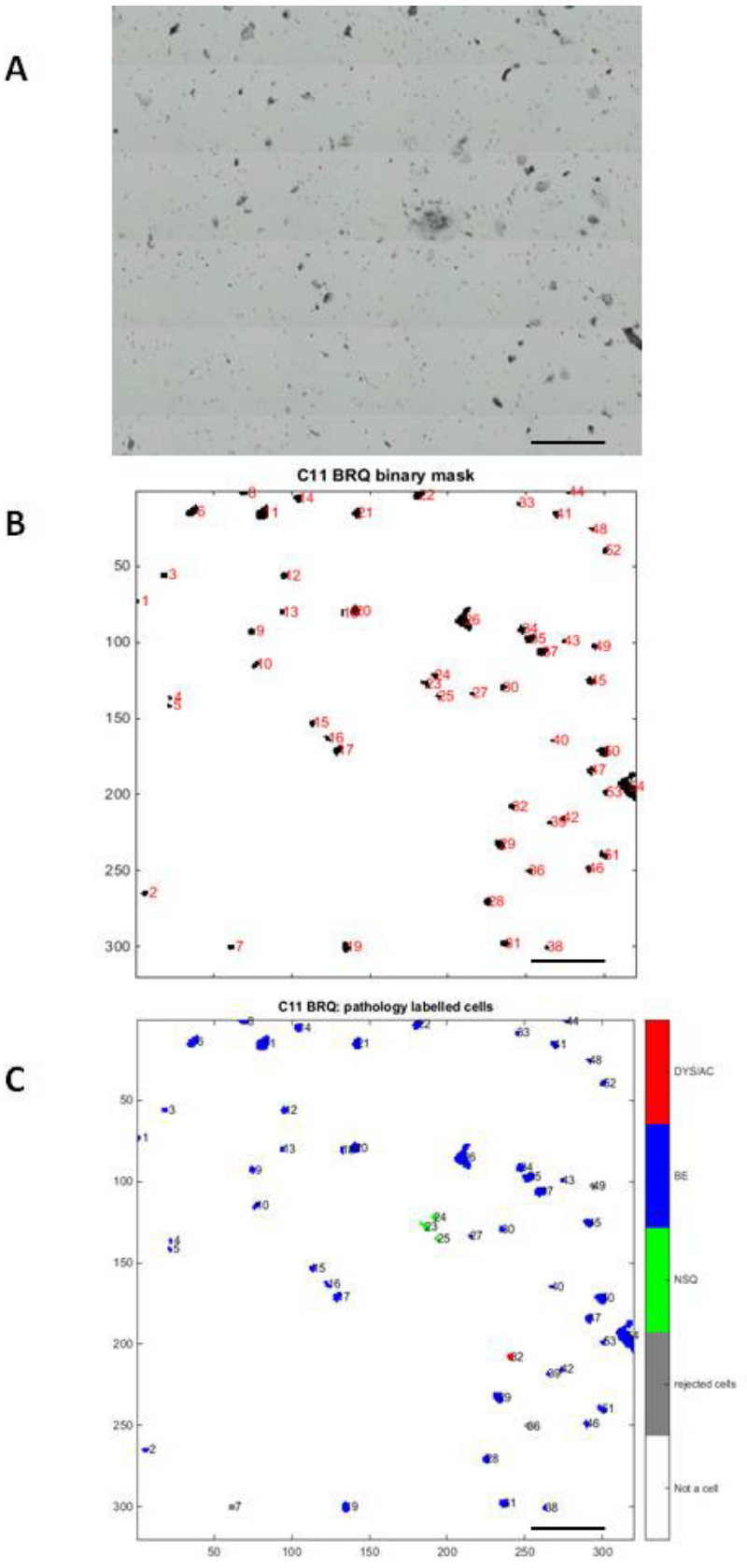


Figure 5-36 Steps involved in assigning pathology labels to every cell A) unstained slide, B) binary mask identifies cell regions, C) training model applied to cells to assign pathology label. The bar in each image represents 300μm.

This greater number of samples partly accounts for the greater number of cell regions identified in the Barrett's and dysplasia/adenocarcinoma groups, but this is also due to the greater cell density (on average) on these slides compared to the normal squamous and stomach samples.

The training model was applied to every cell region identified in the test dataset. A threshold vote system was used as described earlier, using the same threshold of 30% for classification. For most samples, not every cell was classified into the same pathology, so a similar threshold vote was used to classify the whole sample based on the classification of its constituent cells. In this first iteration, a simple majority system was used, whereby over 50% of the cells had to be classified as a certain pathology for the sample to be classified as such. A visual representation of this sequence, from the unstained cells to the pathology labelled binary mask, is shown in Figure 5-36.

The results from application of the training model to this test dataset are shown in Table 5-7.

Overall, the detection of normal squamous and dysplasia/adenocarcinoma samples was reasonably good, with a sensitivity of 79.0% and 83.3% respectively. The detection of Barrett's was very poor, however, with only 31.3% of samples classified correctly. The confusion matrix shows that none of the normal squamous or dysplasia/adenocarcinoma samples was misclassified as Barrett's. The most likely cause for these results is that the training model contained far fewer Barrett's spectra than the other groups. The training model is therefore skewed away from classifying spectra into the Barrett's group, and the Barrett's group can also more readily be affected by a small number of outlying spectra.

Table 5-7 Classification performance of the training model applied to the whole samples as a test dataset. A voting threshold was used for classifying each individual cell on a slide (30% threshold) and for classifying the sample overall (50% threshold). A) Confusion matrix, B) Sensitivity and specificity. NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/ adenocarcinoma.

A

	True pathology classification			
Predicted	**NSQ**	**BE**	**DYS/AC**	**STOMACH**
NSQ	**15**	7	7	7
BE	0	**10**	0	0
DYS/AC	4	15	**35**	14
UNKNOWN	0	1	0	0
TOTALS	19	33	42	21

B

	NSQ	**BE**	**DYS/AC**
Sensitivity	79.0	31.3	83.3
Specificity	77.9	100	54.2

The same argument probably explains the absence of stomach samples classified as Barrett's. Two-thirds of the stomach samples were classified as dysplasia/adenocarcinoma, with the remainder as normal squamous. This suggests that the stomach samples are closer to the glandular cells than the

normal squamous samples, but only limited conclusions can be drawn in the absence of a stomach training dataset.

Only one sample was classified as ‘unknown’, having failed to reach the 50% threshold for classification. Thus for every other sample, over 50% of the cells were classified as a single pathology. This level of homogeneity was not expected given the absence on cytopathology review of clearly glandular or dysplastic cells: many of the Barrett’s and dysplasia/adenocarcinoma samples contained large numbers of squamous cells and cellular debris. For many of these samples, we might therefore expect the majority of cells to be classified as squamous. However this was the case in only 16.7% of the dysplasia/adenocarcinoma samples, and 21.2% of the Barrett’s samples.

There are a number of possible explanations for this. The simplest explanation would be that there are systematic spectral differences amongst cells in the training model that are used for classification, other than true biochemical differences between the pathology groups. This is possible due to the fact that all samples are included in the test set (there is no ‘leave-one-out’ validation), so although the cells used to develop the training model are excluded, spectra from patients included in the training set are also included in the test dataset. Thus patient or sample specific differences could contribute to the result. Nonetheless, since only a minority of the patients were included in the training set (e.g. 5 out of 25 patients in the dysplasia/adenocarcinoma group), such differences cannot explain the result alone.

One possibility is the presence of unidentified dysplastic cells on the slides, although this is unlikely to occur in great numbers. Another possibility is that the biochemical content of the cellular debris is similar to the content of whole cells,

and hence, although such debris may not be useful for classification of cells using conventional cytopathology, the biochemical information is useful for spectral classification. If the cellular debris is simply the result of plasmolysis or fragmentation due to the preparation of the slides, this debris would logically be biochemically representative of the underlying cell pathology.

A further possibility is that biochemical changes precedes a morphological change in cell appearance. If a region of dysplasia or adenocarcinoma underwent a field change (probably reflecting genetic change) that preceded phenotypic change in the cells, there may be a detectable biochemical difference in cells that appear squamous.

To test this idea further, we attempted to develop a training model using only cells that had been classified as appearing squamous on cytopathology, but assigning the cells' pathology to the underlying pathology from which it was taken.

5.13. 'All squamous' training model

As described above, we aimed to test whether cells that appeared squamous, but had been taken from a region of e.g. Barrett's, showed biochemical changes that could be used to identify the underlying pathology.

During the development of the training dataset, many cells that were identified on samples taken from regions of Barrett's or dysplastic histology were classified as appearing squamous on cytology review. We used these cells to build a training model: a summary of the data included (after identical pre-processing steps to the initial training dataset) is shown in Table 5-8.

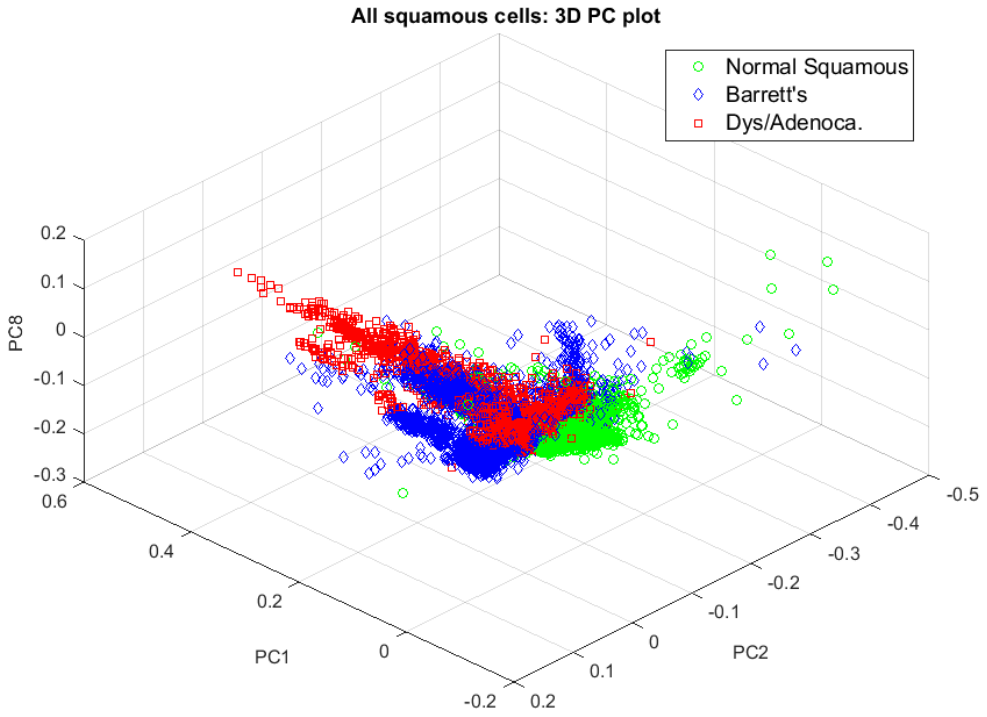
Table 5-8 Summary of data included in the 'all squamous' training model

	No. of patients	No. of cell regions	No. of spectra for analysis
Normal Squamous	8	107	4,714
Barrett's oesophagus	13	148	4,847
Dysplasia/ adenocarcinoma	11	69	2,061
TOTAL	32	324	11,622

With a total of 324 cells and 11,622 spectra in this training set, it is significantly larger than the initial training dataset, which contained 56 cells (the number of spectra changed with each iteration since 22 normal squamous cells were chosen at random).

The training model was developed using the same methods as described earlier, with PCA-fed linear discriminant analysis, followed by LOSOCV. Figure 5-37 shows the spectra plotted according to the most discriminatory PCs (A) and the LD functions (B).

A



B

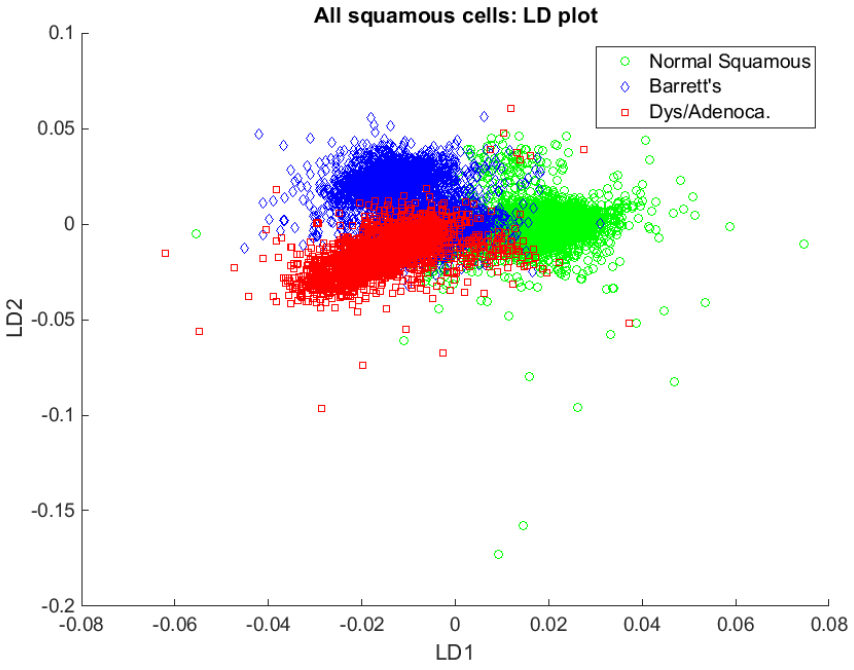


Figure 5-37 Individual spectra plotted according to A) most discriminatory principal components (PCs) and B) linear discriminant (LD) functions, for the 'all squamous' training model.

The PC scatter plot shows marked heterogeneity of the dysplasia/ adenocarcinoma group, with the points separated over a wide distance by PC1. The normal squamous spectra are much more closely grouped, and the Barrett's spectra show some heterogeneity across PC1, but much less than the dysplasia/adenocarcinoma group. It is possible that this reflects the greater heterogeneity within the dysplasia/adenocarcinoma group: this group shows the greatest variation in genetic and histological appearance, and this may be reflected here. Further, the cells in this group could potentially be a mixed population, with some 'normal' squamous cells, whilst others may be abnormal cells showing biochemical changes that precede morphological change.

The results of the 'all squamous' training dataset after LOSOCV are shown in Table 5-9.

The ability to discriminate between the groups is far better than would be expected if the cells from the three different groups (which all appear to be squamous) were biochemically identical. The classification performance is very good for normal squamous spectra, moderate for Barrett's and poor for dysplasia/adenocarcinoma. The poorer performance of the dysplasia/ adenocarcinoma group may result from the heterogeneity implied by the PC scatter plot above. The smaller number of spectra in this group may also contribute to the reduced sensitivity.

The classification of the normal squamous spectra is similar to, but slightly better than, the original training dataset (which had sensitivity 83.6%, specificity 70.6%). This may result from the larger number of spectra in the 'all squamous' training dataset.

Table 5-9 Classification performance of the 'all squamous' training model applied to individual spectra: A) Confusion matrix, B) Sensitivity and specificity. NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/ adenocarcinoma

A)

	True pathology classification		
Predicted	**NSQ**	**BE**	**DYS/AC**
NSQ	**4170**	516	128
BE	408	**3177**	966
DYS/AC	136	1154	**967**
TOTALS	**4714**	**4847**	**2061**

B)

	NSQ	**BE**	**DYS/AC**
Sensitivity	**88.5**	**65.6**	**46.9**
Specificity	**90.7**	**79.7**	**86.5**

Whilst the classification performance overall is too poor to be clinically useful, this does suggest that many of the cells which were classified as squamous by the cytopathologists could have biochemical changes that mark them as different from normal squamous cells. Many of the caveats expressed regarding the results from the original training cells dataset still hold however, and although this dataset is larger, with a total of 32 patients, this remains a small study and conclusions are correspondingly tentative.

5.14. Discussion

The overall aim of this project was to use FTIR to detect Barrett's oesophagus and neoplastic change (dysplasia/adenocarcinoma) in endoscopically collected cell samples. In this study, classification of normal squamous cell samples was performed with sensitivity 79.0%, sensitivity for detecting Barrett's oesophagus was 31.3% and for detection of dysplasia/adenocarcinoma it was 83.3%.

This was based on the application of a training model with the ability to classify individual spectra from the training dataset with sensitivity (after cross-validation) 83.6% for normal squamous cells, 62.8% for Barrett's oesophagus, and 69.5% for dysplasia/adenocarcinoma.

To our knowledge, there were no studies applying infrared spectroscopy to oesophageal cell classification prior to commencement of this work. Previous work using IR to classify cell pathology with similar methodology ('spectral cytopathology', SCP), has shown the ability to classify squamous cell samples from the cervix, urinary tract and head and neck (Bird et al. 2008; Gajjar et al. 2014; Papamarkakis et al. 2010; Schubert et al. 2010), although this work focussed on proof of concept and spectral differences, and did not publish equivalent figures for sensitivity and specificity.

The classification model developed using the training dataset contained small numbers of cell regions in all groups (in the case of the normal squamous group this was deliberately reduced to balance the group size), but this was particularly true for the Barrett's group with only 12 cells. Not only did this have the smallest number of cells, but these tended to be small, isolated cells (rather than a cluster of cells of the same type), and thus contained very few total

spectra (76 in this group, versus 678 in the dysplasia group, and 726 in the normal squamous). This is a very small number upon which to build a spectroscopic training dataset, and this makes the Barrett's dataset prone to strong influence from differences due to individual samples or patients rather than true biochemical differences due to pathology.

Analysis of the inverse second derivative spectra indicated that the key discriminatory peaks seen in the FTIR tissue mapping study, and reported by other studies of FTIR in oesophageal tissue (see Section 4.10.1), were also present in the cells. However, only a few of these features were picked out by principal component analysis, suggesting other differences between the groups were used in the classification model. This is likely to be a major contributing factor in the relatively poor performance of the model when LOSOCV is applied to the training dataset. We might expect this problem to be greatest in the Barrett's group since it is the smallest, and this may partially explain the lowest sensitivity in this group.

Small sample size not only skews the training model towards patient and sample differences, but can also allow a single sample to heavily affect the result. In the case of the dysplasia/adenocarcinoma group, 10 out of 22 training cells were drawn from one sample (and it was noted that this sample stained darker than many others, and this could potentially contribute to 'over-calling' the dark-stained nuclei as dysplastic). The majority of cells from this sample were classified as normal by the training model, and this accounted for all but one of the misclassifications in this group, significantly reducing the sensitivity of the model.

If the Barrett's group in the training dataset has been affected by patient or sample differences, this is a potential explanation for the poor performance of the 2 group model. If based on pathological differences, one would expect the similarities between glandular and dysplastic cells to improve the classification versus normal squamous, as seen in the tissue mapping study. However, if sample differences are introduced to the model which are not based on pathological differences also found in the dysplastic cells, this is likely to impair the performance, as seen in our results.

Given the difficulty in identifying cells of representative pathology on the slides, the application of the training dataset to test samples in which there could potentially be no representative cells of that pathology seems destined to fail. The fact that the training model achieved a reasonably good classification of the dysplasia/adenocarcinoma samples therefore needs to be explained, and a number of possibilities have been discussed above.

Consistent patient/sample differences due to artefact are not sufficient to explain the classification performance of the dysplasia/adenocarcinoma group given the small proportion of patients contributing to the training model. Whilst misclassification of a few cells is possible, this would not have resulted on the homogeneity seen in the classification of these samples, which required over 50% of cells to be classified as dysplastic to be assigned that pathology label.

This suggests that other material on the slide – either cell debris, or cells that appear to have squamous morphology – is contributing to the spectral classification.

To assess whether cells that appeared squamous (and were classified as such at cytopathology review), showed a difference between the pathology groups

we developed a training model based solely on these cells. The Barrett's, and particularly the dysplasia/adenocarcinoma groups, appeared very heterogeneous on PC analysis, which suggests this may contain a mixed population of cells. The ability to discriminate between the squamous cells drawn from the different pathology groups was varied, but considerably better than would be expected if no biochemical difference existed between the groups. This introduces the possibility that cells may exhibit a 'field change', reflecting underlying genetic change, that has led to the histological change observed in that region, but precedes a phenotypic change in appearance in some cells. It is plausible that this cell group may be the most superficial cells, which are also those that are most readily collected by brushing.

Interestingly, a similar finding was reported in the largest study of SCP in cervical samples to date: Gajjar et al. (Gajjar et al. 2014) found that SCP correlated poorly with conventional cytology, but showed better correlation with contemporaneous histology from the same region. This paper cites poor cytology sensitivity and specificity as the reason for this, but an analogous explanation of field change that has not occurred in every cell (and again, particularly in the superficial cells), is another possibility.

This project began as a collaborative project with the group from Northeastern University, to whom we gave a small number of samples for analysis; this work has recently been published (Townsend et al. 2015).

Based on a total of 10 samples, the authors developed 2 group classification models comparing (in turn) normal squamous, Barrett's and dysplastic cells, using a random selection of 400 spectra in each group to balance the group size. They report a sensitivity to detect Barrett's versus normal squamous of

95.5%, normal squamous versus dysplasia 93.4%, and Barrett's versus dysplasia 88.7%. However spectra from the same patients were included in both test and training datasets, and since these results are drawn from a very small number of patients it is possible that 'between patient' differences (as opposed to 'between pathology' differences) contributed to this result. Also, since both the analysis by Townsend et al. and the present study rest on very small numbers of patients, there is potentially a wide margin of error in both studies. Of note however, the cell preparation protocol in this study did not seem to encounter the same level of difficulty that our study found in identifying representative cell samples: this study included 145 Barrett's cells and 14 dysplastic cells from a total of 10 samples. This study used a different staining protocol from ours (Papanicolau staining as opposed to 'Diff-Quik'), and used xylene and a cover slip to prevent degradation after staining – these factors may have improved cell staining and preservation and aided cell identification.

Further, the discriminatory spectral features seen in the work by Townsend et al. are similar to the findings seen in the inverse second derivative spectra in our study (it should be noted that Townsend et al. used second derivative spectra as the basis for their classification model), and also correlate with the findings from our tissue mapping study and previous FTIR work in the oesophagus. For example, the amide I peak at around 1650cm^{-1} is strongest in the normal squamous group, the DNA peak at 1235cm^{-1} strongest in the dysplastic cells, and the glycogen peak at 1020cm^{-1} is strongest in the normal squamous cells and almost absent in the dysplasia group. This strengthens the findings in their study, and lends further weight to an argument for an underlying biochemical difference between the pathology groups, as seen in our analysis of the inverse second derivative spectra.

A further finding in the study by Townsend et al. was that cells which appeared squamous classified according to the underlying tissue histology. This supports our finding that squamous cells from the different groups show spectral differences which may be used for classification, and the findings of Gajjar et al. This consistent finding therefore supports the theory that genetic and biochemical change may precede morphological change in some cells.

However, others using an alternative approach have argued that cytological changes precede histological changes. Reid et al. used flow cytometry in the oesophagus and identified some patients with IM only on biopsy who had cytological abnormalities which suggested dysplasia (Reid et al. 1987). They argued that cytology could therefore be used as an early predictor of dysplastic change prior to histological evidence of dysplasia. A very similar finding was made by Borovicka et al. (Borovicka et al. 2009): in a large study using digital cytometry, they identified a small subgroup of patients with more advanced cytology than histology. Of this subgroup, 21% progressed to more advanced histological disease on follow up.

These arguments may be two sides of the same coin: an explanation for both effects would be that genetic change occurs and spreads throughout a cell population in a given tissue region, and cells then variably express the new phenotype. Early in the process, a small population of cells may be detected with dysplastic features that precede histological change. Once histological change occurs, there may remain a population of cells that have not yet undergone phenotypic change.

As has been mentioned, the small number of patients, cells and spectra in the training dataset, and especially the Barrett's group, has been a major limitation

of this study. The factors which contributed to the small sample size, along with other limitations of the study, are outlined in section 5.15 below. Potential further work based on these findings and their future application to clinical practice are discussed in the Summative Discussion (Chapter 10).

5.15. Limitations of the present study

There are numerous limitations to this study or areas which could be improved in future work. A crucial element of this study was the development of a reliable training dataset based on representative cells of known pathology. The development of this dataset was the key area of weakness of the study: many steps were involved in transferring cells from the patient to the calcium fluoride slide for subsequent measurement, staining and cytology review, and many of these steps could be improved. Since this study began, a paper has been published that summarises the optimal steps in preparing cells for infrared spectroscopy (Baker et al. 2014), which is referenced below and would be helpful in guiding further work.

The number of cells collected at endoscopic brushing appeared to vary enormously based on the apparent density of cell pellets at slide preparation. The brush used was the same in every patient, but the technique of performing the brushing was not perfectly standardised. Further, the initial plating experiment seemed to show more samples were affected by heavy bloodstaining than had insufficient cellularity. We therefore attempted to brush less vigorously to reduce the bloodstaining. However in retrospect, these early samples were mostly dysplasia/adenocarcinoma samples which in general had

much higher cellularity and a greater tendency to bleed, and the less vigorous brushing probably contributed to the reduced cellularity of some later samples.

A further problem with collection was the high presence of cells that appeared squamous in samples taken from regions with glandular histology, as has been discussed. It is possible that this was exacerbated by lighter brushing, which is more likely to sample only the most superficial cells, which may include squamous cells that originated more proximally in the oesophagus that have sloughed off and been deposited more distally over a glandular region.

Differences attributable to the choice of fixative (CytoLyt™ versus formalin) used to store the cells have been discussed above. For future work, in terms of spectral contributions the most important factor would be to use only one fixative throughout, although CytoLyt™ is probably better since it limits the artefact in the carbohydrate region between 1000-1080cm^{-1}, and would also be preferred on the basis of reducing cell clumping and the presence of red blood cells.

A more general problem with fixative is ensuring maximal cell survival, and particularly in preserving the glandular cells, since any differential preservation would contribute to the paucity of glandular cells.

Prior to plating, some groups advocate washing cells in Hank's Balanced Salt Solution (HBSS), rather than NaCl as we used, as HBSS removes phosphate ions before cytospinning (Baker et al. 2014). The presence of excess phosphate ions could lead to overestimation of phosphate (and associated peaks, such as those frequently assigned to DNA due to phosphate content) on spectroscopy.

A further source of error in plating could have been the centrifugation speed. We used 1500rpm for 3 minutes in accordance with current standard protocol in our cytology department (this equates to approximately 2,100*g*). However, Baker et al. advise using a lower force of 800*g* for 5 minutes, and warn against the problem of small cells 'bouncing off' the slide, in which case they recommend a protocol of 400*g* for 5 minutes, followed by 800*g* for 5 minutes. It seems very likely that this contributed to the very low number of (smaller) glandular cells seen on our slides.

Other methodological steps that may have contributed to the difficulties in identifying cells were the staining and photographing of cells. Townsend et al. (2015) used a different staining technique (Papanicolau) that may have been superior to our 'Diff-quick' technique. They also used xylene and a cover slip to prevent degradation between cell staining and review. It seems likely that a small number of cells washed off the slides during staining, contributing further to cell loss.

Photographing the cells for review introduced subjectivity, and the possibility of missing cells. Although I have received some training in cytology assessment I have very limited experience compared to a cytopathologist or cytology technician. Attempting to compensate for this by photographing any cell that appeared to be even possibly glandular may not have been sufficient. The other great difficulty with photographing the cells was correlating positions with the binary mask. Although for some samples this was relatively straightforward, in samples with heavy cell density, or where cells had washed off in staining, this was very challenging.

As described for histology in the tissue mapping sections, cytopathology review is a potential source of error. In this study it was performed by two cytopathologists together, not independently. Further cytopathologists are not used to reviewing single cells in isolation, but more usually look at a whole sample for assessment. Presentation of cells in isolation could mean the decision is affected by variable staining between slides, which might be accounted for by taking the slide as a whole.

Comparison with cervical cytology can give some indication of the performance accuracy of conventional cytology review. Cervical cytology is performed in high volume by highly trained and experienced cytotechnologists, however one systematic review found sensitivity ranging from 30-87% (Nanda et al. 2000). In one large study in England, of 3,759 slides reported as normal, only 45% were subsequently deemed to be normal on a second review, and 23% were found to have moderate dyskaryosis or worse (Castanon et al. 2012). Since cervical cytopathology is performed in much higher volume than oesophageal cytopathology, it might be expected that the accuracy of oesophageal cytopathology is less good.

In summary, many factors contributed to the low numbers of glandular cells available for the training dataset. This may not be a problem unique to this study: low numbers of Barrett's cells per sample were also a problem in BEST2, as the authors of this study report that some of the weakness in sensitivity was due to Barrett's patients not having any glandular cells in the sample (Ross-Innes et al. 2015).

The other potential source of error in this study comes from data processing.

Pre-processing steps could remove useful spectra and diminish true variation between groups. Alternatively, the inclusion of anomalous spectra (particularly 'edge' pixels) could also impair classification. Mie scattering correction runs the risk of 'overfitting', using curves that can fit any dataset, and therefore the process has a minimal effect.

The major processing step in this study was the binary mask used to process maps and identify cells. This used a combination of spectral information (the 1650cm^{-1} peak) and size criteria to identify cells. The use of size criteria in the binary mask is affected by cell clumping, and may exclude large numbers of cells if clumped together.

The spectral information is non-specific, and will identify any protein signal that gives a large amide I peak in this region. It can also miss cells with a relatively weaker signal at 1650cm^{-1} if they have strong signal elsewhere in the spectrum. Other spectral markers could be used, either alone or in combination, to more accurately identify cells. One approach could use a DNA marker (e.g. the 1234cm^{-1} peak) to identify cell nuclei, in combination with associated cell cytoplasm (represented by the 1650cm^{-1} peak). This could be used in combination with size criteria to quantify the amount of nuclear material present.

There is much potential to refine the binary mask with a combination of further spectral information and size criteria to give a highly accurate cell identification tool.

Alternative approaches could have been used to overcome some of the problems with small numbers of cells and identifying individual cell types. If a simpler approach using measurement of cell pellets had been used, this would have ensured large numbers of cells were measured and a stronger infrared

signal. However this would have introduced difficulties with multiple cell types being present in each measurement.

Alternatively, cultured cell lines could have been used to provide homogeneous cells from which to measure each group (i.e. squamous cells, glandular cells and dysplastic/cancerous cells). This would have provided a measure of the intra-group variation and a reliable cell population from which to identify the differences between groups. Knowledge of these differences could then have been applied to patient samples.

SECTION C: TOWARDS CLINICAL TRANSLATION OF RAMAN SPECTROSCOPY FOR DIAGNOSIS OF BARRETT'S-ASSOCIATED NEOPLASIA

CHAPTER 6: INTRODUCTION AND OBJECTIVES

6. Introduction

Endoscopic surveillance for Barrett's oesophagus depends on accurate identification of dysplasia and adenocarcinoma to achieve the goal of early detection and allow timely intervention. These neoplastic changes are currently identified using endoscopy and biopsy. Current guidelines recognise that areas of neoplasia may be missed at conventional white-light endoscopy, and consequently in addition to biopsy of any visible lesions, a random biopsy protocol is recommended, taking quadrantic biopsies every 2cm in the columnar segment (Watson et al. 2005). Nonetheless, since this only samples a tiny fraction of the Barrett's mucosa, areas of neoplasia may be missed. Many advanced endoscopic imaging technologies have been developed to improve detection of dysplasia, with wide ranging reported sensitivity and specificity. Of these the most widely used is narrow-band imaging, but none has demonstrated sufficient superiority over white-light endoscopy to become universally used in clinical practice. Consequently, there remains a clinical need for improved detection tools in Barrett's associated neoplasia.

Once biopsies have been taken for Barrett's surveillance, they must be processed and analysed by a pathologist. Inter-observer variability in reporting Barrett's dysplasia is well recognised and discussed above. This is therefore a further potential source of error for patients undergoing Barrett's surveillance, with potentially significant treatment consequences. A tool that could provide a more standardised classification, based on an objective measure such as tissue

biochemistry, is therefore highly desirable. If such a tool could be applied *in vivo*, this would allow real-time classification of pathology.

In recent years there have been major advances in endoscopic treatments for HGD and early cancer, allowing patients to avoid radical surgery in the form of oesophagectomy, with its associated morbidity and risk of mortality. This not only reduces the risks for patients who would previously have undergone oesophagectomy, but offers the possibility of curative treatment for many more patients who would not have been considered fit enough to survive major surgery. Whilst the modes of delivery vary between therapies, in order to deliver these treatments accurately and achieve oncological eradication, regions of dysplasia or early cancer must be delineated with precision. Consequently this shift in the treatment paradigm towards endotherapy presents a pressing need for real-time endoscopic identification of neoplasia in the oesophagus.

Raman spectroscopy is an optical spectroscopy technique that has shown considerable promise as a diagnostic tool based on its analysis of tissue biochemistry. Raman spectroscopy has a number of characteristics that make it an attractive candidate for real-time endoscopic imaging in the oesophagus: it provides rapid, objective, non-invasive tissue diagnosis. By providing a qualitative and quantitative measure of the biochemical changes occurring in the sample, such as proteins, nucleic acids and glycogen content, disease can be detected at an early stage. It may be possible to identify the first steps of neoplasia occurring even before the changes in tissue architecture that can be classified at histology.

Previous work with Raman microspectrometers has demonstrated the ability to discriminate distinct pathologies using *ex vivo* oesophageal tissue specimens, including Barrett's metaplasia, LGD, HGD and adenocarcinoma (Kendall et al.

2003; Shetty et al. 2006). Recent evaluation of a custom-built fibre-optic Raman probe (designed to function with a conventional medical endoscope) has shown the ability to identify HGD and early cancer in a clinically applicable timeframe with sensitivity 86% and specificity 88% (Almond et al. 2013). These results were based on the reuse of a single in-house manufactured probe, and further work is required in order to translate these highly promising *ex vivo* results to a larger scale *in vivo* study. A key step in developing the probe for clinical use is to demonstrate consistency and reliable performance between multiple probes of the same design.

Previous work investigating the biochemical basis of the differences detected by Raman and used to classify pathology has been performed using point spectroscopy, or maps with low (~100μm) spatial resolution (Almond et al. 2013; Bergholt et al. 2014; Hutchings et al. 2010; Kendall 2002; Shetty et al. 2006). This work has tentatively concluded that normal squamous tissue contains a higher glycogen content than metaplastic/neoplastic tissue, and metaplastic/neoplastic tissue a higher protein and nucleic acid content. Recent advances in Raman mapping instrumentation and computing power have enabled large datasets to be acquired using high quality Raman mapping with microspectrometers capable of high resolution (1.1μm). Application of these techniques in the oesophagus could investigate further the biochemical basis of Raman in discriminating between pathology groups in the oesophagus.

Previous development of the Raman probe in our group has been done with an excitation wavelength of 830nm, however other groups have used 785nm. Whilst the Raman signal should be independent of excitation wavelength, the wavelength will affect the signal strength and the fluorescence background, and thus optimising this parameter would enhance performance of a Raman probe.

This issue is not limited to the probe: Raman mapping work has been reported at a range of excitation wavelengths. The ability to combine data from different spectrometers, such as applying a training dataset from one Raman system to a test dataset measured on another, is a key requirement if Raman spectroscopy is to be used in clinical practice more widely, but there is difficulty in comparing results between spectrometers when measured at different excitation wavelength. Whilst some have attempted to address this problem for simple organic compounds with a calibration protocol (Bocklitz et al. 2015), this is likely to be more challenging for complex biological samples. If a certain excitation wavelength were known to be optimal for oesophageal tissue then this could be used as standard, more readily allowing comparisons of data collected on different Raman systems.

For mapping applications, this issue is not limited to wavelength but also modality: both infrared and Raman spectroscopy may be used. Infrared and Raman spectroscopy have a number of differences (summarised in Table 2-3 in the literature review), and each has its own strengths and weaknesses. There may be a number of reasons for choosing one modality over another, but where the ultimate goal is identification of pathology, the sensitivity and specificity of classification are clearly of great importance. Quantifying the difference between modalities with regard to classification accuracy would therefore be valuable, and inform decisions when choosing the optimal technique for tissue investigation.

6.1. Objectives

1. Translate Raman probes from *ex vivo* tissue measurements to first *in vivo* measurements.
    a. Evaluate 5 prototype fibre-optic Raman probes built to a novel design, prior to larger-scale manufacture for *in vivo* use.
    b. Optimise parameters for *in vivo* data acquisition.
    c. Initial safety and feasibility testing of probe in first patients *in vivo*.
2. Explore the molecular basis of Raman spectral differences that underpin the classification of pathology, using Raman mapping at high resolution with a new generation micro-spectrometer with long acquisition times.
3. Compare classification performance of Raman maps measured using 785nm, 830nm and FTIR maps of the same region of tissue (the work presented above in the FTIR mapping section formed the FTIR component of this comparison).

6.2. Sample collection

Samples for all 3 arms of this study were collected using the same ethical approval, consent form and methodology. All tissue samples were then used as a tissue bank from which to select samples for each individual arm of the project.

A summary of all tissue samples included in the study, with outcome of double pathology review, can be found in Appendix VII.

6.2.1. Ethical approval and informed consent

All patients participating in the study provided informed consent. Each patient received a standardised information leaflet explaining the study (Appendix V). Patients had time to read the leaflet prior to a face-to-face discussion with the researcher, which offered an opportunity to ask any questions about the study. If happy to proceed, patients then provided their consent to participate, and signed a standardised consent form (Appendix IV). No samples were taken from patients unless they had provided informed consent.

Ethical approval for the study was obtained from the Gloucestershire Local Research Ethics Committee (Appendix VI).

6.2.2. Sample collection

Samples were collected from patients undergoing scheduled endoscopy, either as part of Barrett's surveillance or indicated for other clinical grounds. In order to enrich the sample population for patients with oesophageal adenocarcinoma, some samples were also collected from patients undergoing surgery for oesophageal cancer.

Tissue samples were collected at endoscopy using endoscopic tissue forceps (Olympus Medical, Tokyo, Japan) passed down the instrument channel of an endoscope, under direct vision at endoscopy, or at endoscopic resection using diathermy.

The sample was then placed into a Cryovial (Thermo Scientific) and immediately snap frozen in liquid nitrogen, before storage at -80°C.

For the samples collected at surgery for oesophageal cancer, once the oesophagus was removed from the patient it was opened longitudinally, washed in distilled water, and samples were collected using a scalpel. Samples were placed in Cryovials and snap frozen as above.

6.2.3. Sample preparation and measurement

Sample preparation, histology consensus reporting, and selection of regions for mapping were the same for all mapping measurements, and using identical methods to those described for FTIR mapping above in sections 4.3-4.4.

CHAPTER 7: EVALUATION OF RAMAN PROBES

7. Evaluation of Raman probes

7.1. Probe design

The Raman probes evaluated in this project were designed by the Raman for Early Cancer Diagnosis (RECaD) consortium (Gloucestershire Royal Hospital, University of Exeter, University of Bristol) and built by the Interface Analysis Centre at the University of Bristol, and PA Consulting, Cambridge. An earlier probe design has been described in detail in a previous publication (Day et al. 2009). A schematic of this earlier probe design is shown in section 2.2.2.2 in Figure 2-5. The current probe design has undergone a number of modifications, but is conceptually analogous to the schematic in Figure 7-1. The exact internal layout of the probe is awaiting patent and therefore not illustrated here. A photo of a packaged probe is shown in Figure 7-2.

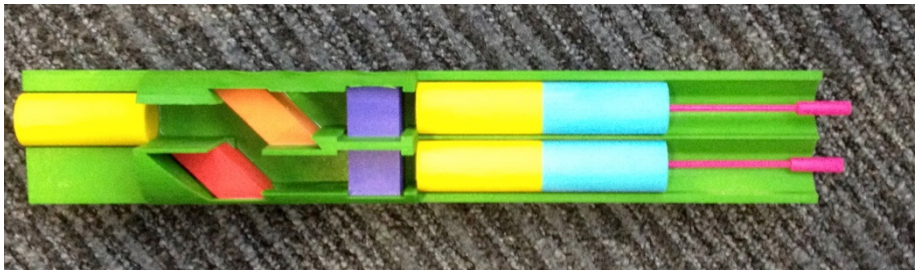


Figure 7-1 Mock up of probe design using coloured 3D printer. Green is the probe template/scaffold and the coloured components are the optics. Each tubular lens and ferrule is 1mm in diameter and the total probe has a diameter of <2.8mm when encapsulated in a stainless steel sheath.

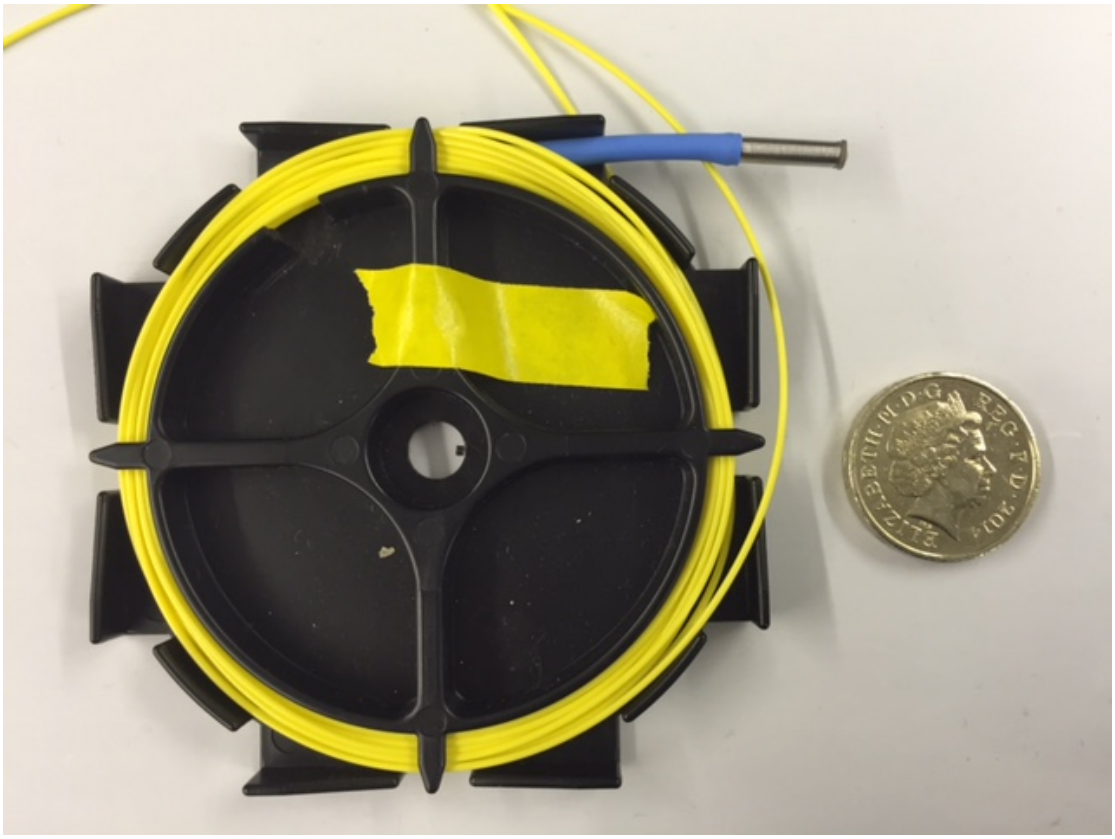


Figure 7-2 Packaged Raman probe with one pound coin for size reference.

Laser light enters the probe via the excitation fibre, where it is collimated by a Gradient Index (GRIN) lens, and filtered to remove Raman signal and elastic scattering generated within the fibre optic excitation fibre. The light then travels through a filter which only allows light of specific excitation wavelength (830nm) to pass through, and onto a GRIN lens to be focussed onto tissue. The mean collection depth in tissue is 150μm, with the aim of allowing contact with tissue and analysing the epithelium. Raman scattered light from the tissue is collected by the GRIN lens and collimated. The light then passes to a filter which removes elastically scattered light and allows inelastically scattered light to pass through. The light is then focussed onto the collection fibre by another GRIN lens for return to the spectrometer.

The housing of the probe is 3D printed from a polymer, to allow accurate, reproducible and low cost manufacture. The internal surface of the probe is gold coated which reduces background signal from the polymer housing, and makes assembly easier as it provides contrast to see the small component parts.

The Raman probe is connected to a laser source and spectrometer as shown below in Figure 7-3. The excitation fibre of the probe is connected to a monochromatic laser at 830nm wavelength. The collection fibre carries light from the probe to a Renishaw System 100 spectrometer (Renishaw plc., Wotton-under-Edge, Gloucestershire). Light is then filtered and collimated, then reflected off a dispersion grating and focussed onto a charge-coupled device (CCD) detector.

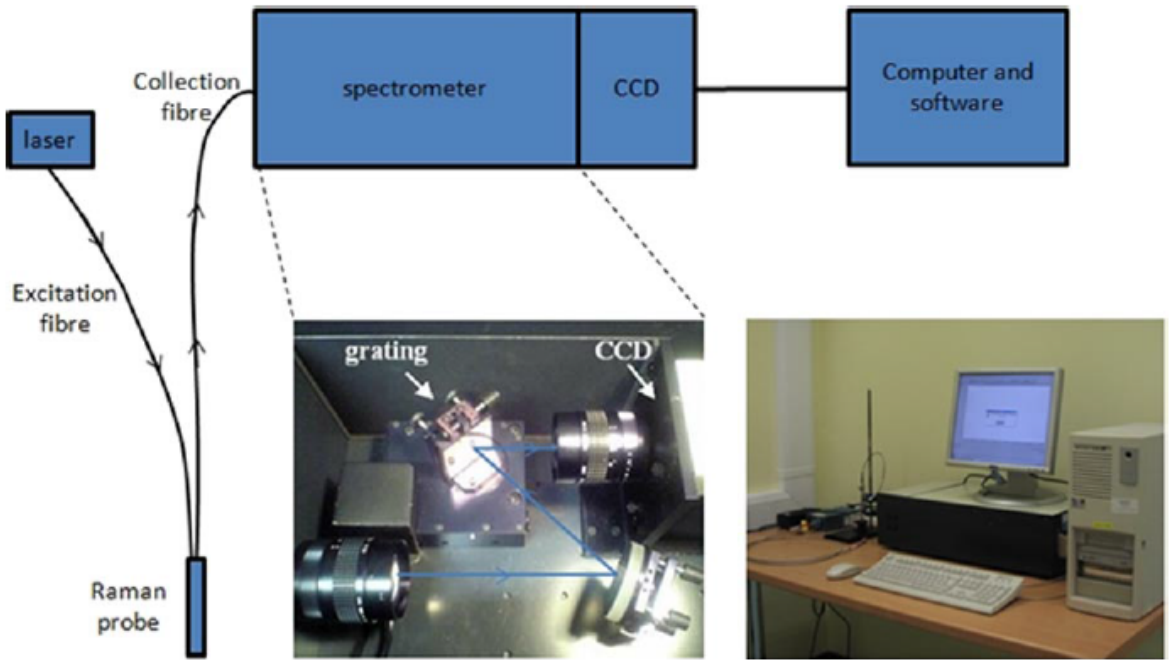


Figure 7-3 Schematic of laser, Raman probe and spectrometer, showing internal layout of the spectrometer. Reproduced from Almond et al. 2011.

7.2. Prototype Raman probe testing

Unfortunately, due to problems identified on the initial probe evaluation it was not possible to perform the programme of experimental work originally intended. The experimental plans are outlined below, and the problems discovered during initial testing are discussed in section 7.2.1. The full list of planned experiments is outlined in more detail in Appendix IX.

The initial phase of the evaluation aimed to test reproducibility of the probe using measurements of known reference standards using multiple probes, spectrometers and operators. The effect of the incident angle of light was to be evaluated. Probe stability over time was to be assessed. The depth of measurement was to be accurately determined using reference standards placed at varying depths.

Once these assessments were complete using measurements of reference standards, tissue measurement would begin. Reproducibility of probe measurement on tissue, along with the effects of repeated measurement on tissue, would then be assessed. Other effects which may be present during endoscopic measurements *in vivo* would then be tested, including the effects of saline, water and anti-foam lavage on measurement, and the presence of endoscopic white light and narrow band imaging.

Once the reproducibility testing was complete, assessment of the diagnostic performance of the probes would begin, with *ex vivo* tissue measurement. Oesophagectomy specimens would be collected in theatre, and measurements made on fresh tissue. After opening the specimen and cleaning with saline, a grid would be placed on the tissue surface to ensure consistency between spectral measurement and biopsy position. At each position on the grid spectral

measurements would be taken using a range of acquisition times (0.1s, 0.5s, 1.0s) followed by a biopsy from the same point. This would enable correlation of the histopathology report for a given tissue sample with its tissue spectrum, and development of a training set of data that could then be used to build a classification model to predict histology from the Raman spectrum.

In addition to measurements on fresh *ex vivo* tissue samples (which would necessarily be from patients undergoing oesophagectomy for either adenocarcinoma or less commonly squamous cell carcinoma), snap frozen tissue from biopsies and endoscopic resections would be used to provide more samples of Barrett's oesophagus and dysplasia as opposed to cancer. These tissue samples would be fully defrosted prior to spectral measurement. The number of measurements and biopsies would depend on the size of the tissue section.

In addition to developing a classification model, use of measurements across the oesophagus with the grid could also be used to gain an estimate of how many biopsies/probe measurements would be required per unit area to give a certain sensitivity for detecting pathology. This would be useful in planning how the probe would be used *in vivo*, and determining the number of measurements required for surveillance if the probe was to be used as a single modality (in addition to white light endoscopy).

Once the reproducibility and stability of the probes had been determined we aimed to apply for ethical approval to perform initial safety and feasibility testing *in vivo*.

7.2.1. Initial probe testing and loss of power output

The initial batch of 5 prototype probes was assembled by collaborators from the University of Bristol and PA Consulting in Cambridge. On the day of assembly, each probe was tested by passing the laser light (830nm) through the probe for several seconds and measuring the laser output. Each of the probes were found to give a high laser output (approximately 100mW). Following an uneventful transport to Bristol the probes were then tested in Bristol two weeks later. Immediately after the laser was turned on there was a high laser output, however this began to drop off after 2-3 seconds, and then rapidly declined in about 10-15 seconds to a low level (<10mW). This was not expected and suggested that there was serious problem with the transmission of the laser beam through the probe.

7.2.2. Inspecting the probe components

The fibre-optics were checked and found to be intact, transmitting high laser output. After disassembling the probe it was discovered that both of the gradient index (GRIN) lenses had become covered in a metallic/brown residue (see Figure 7-4).

A

B

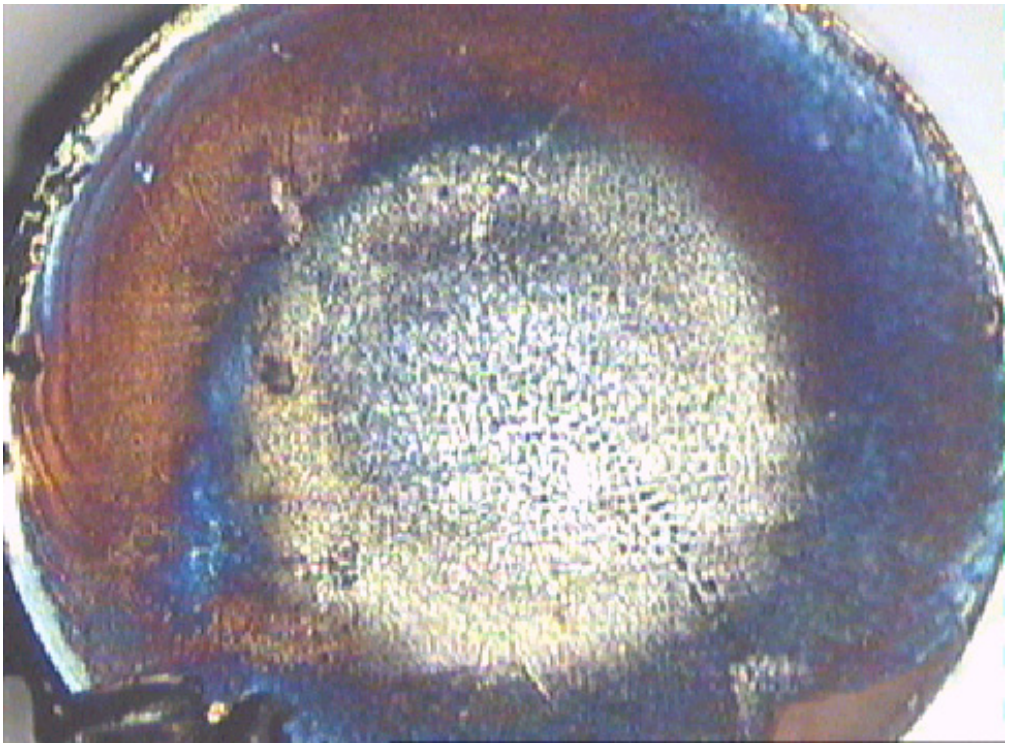


Figure 7-4 A) Gradient Index (GRIN) lens at laser entry point, B) GRIN lens covered in residue at laser exit point.

The pattern of the residue on the laser input side shows greatest residue density around the spot of laser input, then a somewhat haphazard distribution around the surface elsewhere. The pattern of residue density on the laser output lens (at which point the laser should be a collimated beam) shows an even distribution closely corresponding to the size of the laser beam, with a reduction in density of the residue with increasing distance from the laser beam.

The pattern appears to correspond to a vaporised deposition. This fits with the observed pattern of loss of power, with the hypothesized mechanism as follows: as the lenses heat up (100mW laser will cause heating) this could cause vaporisation of a substance within the probe casing which then deposits onto the hot GRIN lens. As the lens darkens so it will absorb more energy and heat up faster, causing further vaporisation and deposition on the lens. This deposition will reduce the laser power output from the probe.

7.2.3. Spectral analysis of residue on the lenses

A number of possible substances were considered as candidates that could have been deposited onto the lens: the gold used to coat the inside of the probe casing, the polymer used as substrate for 3D printing the probe casing, Epoxy resin used as part of the casing, or the glues used to hold the components in place. Raman spectra were measured from each of these substances, and compared against the spectra from the deposits on the GRIN lenses.

The Raman spectra from the GRIN lenses were measured, and slides were prepared of the various resins used, to allow spectral comparison. The resins

were vaporized onto glass cover slips (Loctite UV adhesive, Norland UV adhesive, 2 part epoxy resin, and the polymer used as 3D printing substrate). The Raman spectrum from gold was not tested: since it has no molecular bonds, pure gold does not have a Raman signal. Raman spectra were then measured from each of the resins in turn, using a Raman microspectrometer. The spectra from these measured resins, plus the spectra from the deposition on the GRIN lens, are shown in Figure 7-5.

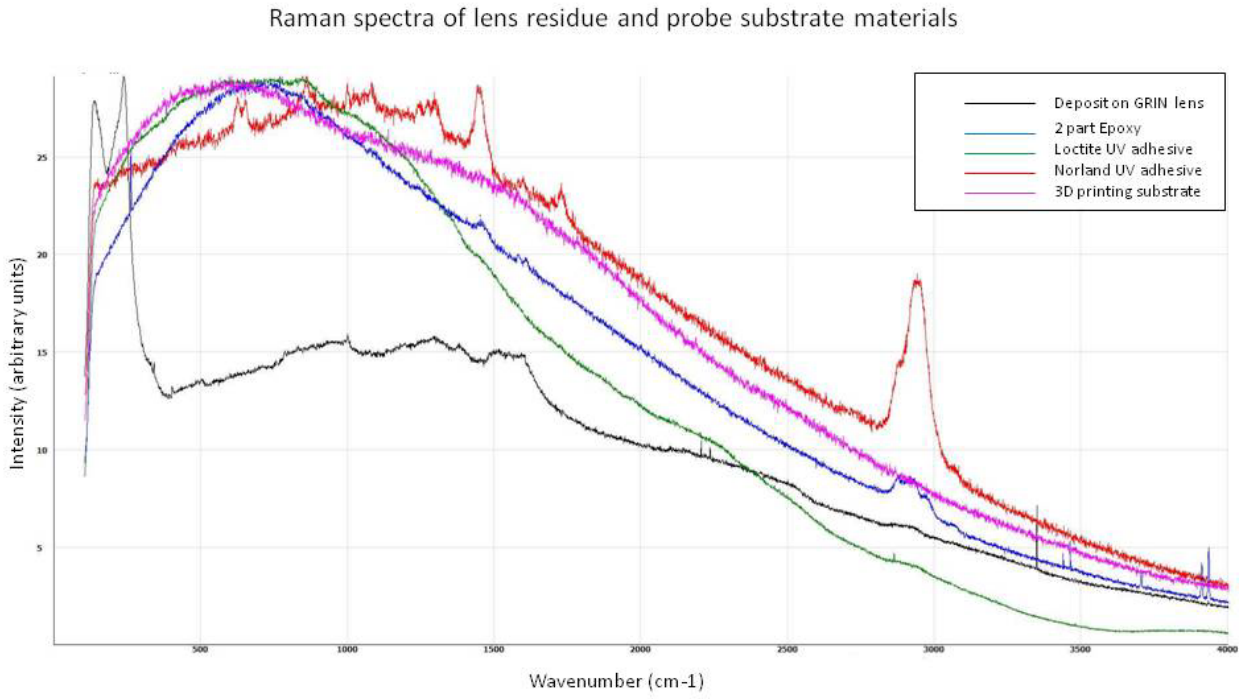


Figure 7-5 Raman spectra of lens residue and probe substrate materials

The residue deposited on the GRIN lenses showed major peaks at around 200cm^{-1} and 250cm^{-1}. Corresponding low wavenumber peaks were not seen in any of the resins tested, suggesting the major component of the residue was not any of these resins. The GRIN spectra also showed a very small peak at 1000cm^{-1} (corresponding to phenylalanine), and small peaks around 1250-

1300cm^{-1}, which were also present in the Norland UV adhesive. There is a large peak in the Norland UV adhesive at around 2900cm^{-1}, with a very small peak in the GRIN lens spectrum at the same wavenumber. This peak is also present in the 2 part epoxy spectrum.

Whilst it is possible that some of the resins (particularly the Norland UV adhesive) may have contributed to the deposit on the lens, the Raman spectra do not closely correspond with any of these substrates. Large signals at low wave numbers are often seen with metal oxide compounds, but given the chemical properties of gold, it seems unlikely that gold oxide is the chief source of the deposit.

The most plausible source of the deposit seemed to be a combination of the housing polymer with the gold coating, but the Raman analysis of the spectra have not confirmed this.

7.3. Discussion

The deposits which formed on the GRIN lenses reduced laser output power and hence prevented the planned programme of experimental work from being completed. The most likely source of the deposits was the polymer used in 3D printing and the gold coating used in manufacturing to coat the interior surface of the 3D printed probe body. Clearly this caused a significant problem and must be addressed in future iterations of probe design. In order to ensure that future designs do not encounter this problem, future work is planned by our collaborators at the University of Bristol to build up a model probe sequentially, starting with the 3D printed body, a fibre optic cable, and a GRIN lens, then gradually add in one component at a time. Sequential testing in this way should

identify any other sources of vapour deposition that could potentially coat the GRIN lens and lead to the same problem in future.

Further funding has now been secured to support this development work and manufacture new probes. Once successful probes are built, the package of planned experimental work could then be undertaken to assess the reproducibility and stability of the probes, then proceeding to evaluate their diagnostic capability.

A reproducible, stable, and diagnostically sensitive and specific probe would have great clinical value. Success in optimising the Raman probe for *in vivo* use would provide the necessary evidence to inform a clinical trial in patients to evaluate the Raman probe. The planned clinical trial could then provide supporting evidence to implement this technology widely to allow patients to benefit.

Potential applications and clinical implications of this technology are discussed further in the summative discussion (chapter 10).

CHAPTER 8: RAMAN MAPPING TO EXPLORE THE MOLECULAR BASIS OF RAMAN CLASSIFICATION

8. Raman mapping to explore the basis of Raman classification

8.1. Instrumentation

Raman maps were measured using a new prototype Renishaw 802 micro-spectrometer at the Biophotonics Research Unit in Gloucestershire Royal Hospital. This instrument is still in development and is not yet generally available.

The principles of measuring Raman spectra are outlined earlier in the thesis. Raman mapping is performed by moving the sample on a motorised stage under the laser beam. Renishaw's Streamline^{TM} technique uses a linear laser beam, which passes across the sample as the stage moves, creating a series of parallel spectra on the detector. For a given point, the spectrum moves across the detector as the stage moves, and signal is collected on the detector continuously as the laser line raster scans across the sample (see Figure 8-1).

The 802 system uses excitation wavelength 785nm, and uses a Leica long working distance 50x objective lens to focus the laser beam.

Prior to sample measurement the instrument must be calibrated. The 802 performs an automated calibration procedure known as 'Performance Qualification' to optimise the system alignment and make wavenumber adjustments using known reference standards.

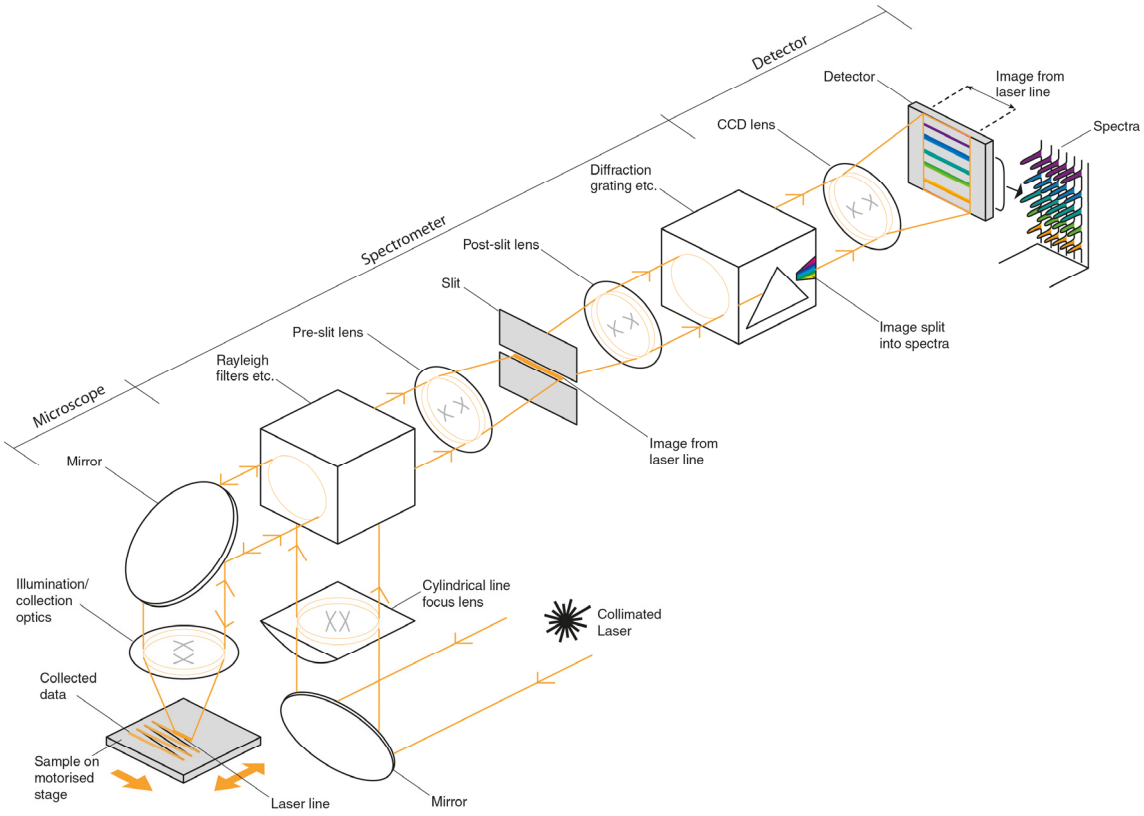


Figure 8-1 Schematic diagram of a Raman microspectrometer using a Streamline™ laser beam to raster scan across a sample to create a Raman map. (Reproduced from http://www.renishaw.com/en/streamline-generate-chemical-images-rapidly--9449)

Calibration is essential to ensure the CCD detector records consistent wavenumber positions to provide reproducible readings and allow comparison of spectra measured at different times.

Silicon is used to calibrate the Raman shift wavenumber value as it has a single sharp peak at 520.4cm^{-1} which can be used as a reference. If a peak shift has occurred in the system and the recorded wavenumber for this peak is incorrect, a wavenumber offset can be applied to calibrate the system accordingly, bringing the measured value back to 520.4cm^{-1}.

Neon-Argon emission lamps are used as they provide an accurate source for a number of atomic emission lines across the absolute wavenumber range 10000 to 12100cm^{-1}. The 802 has an in-built neon-argon lamp, and calibration aims to fit the measured spectra to the absolute peaks with the minimum error through optimising system alignment.

A polymer measurement is also recorded to ensure peaks are measured across the entire spectral range.

8.2. Sample selection and measurement

The samples used for this study were selected on the basis of single pathology review, prior to review by a second pathologist. The tissue classification used was therefore based on a single opinion only (unlike the other tissue mapping studies in this project). A small number of the samples were classified differently by the second pathologist – this issue is considered further in the discussion.

The tissue section on calcium fluoride was defrosted prior to measurement. For this study, high quality spectra with relatively long acquisition times from regions of homogeneous pathology were required. These regions were identified as described above for FTIR mapping, and the region of interest was identified on the unstained tissue section using the microscope. Raman maps were collected using 1.1μm step size, and an acquisition time of 60s. To ensure homogeneous pathology and to reduce the length of time taken to measure each map, the majority of maps were measured across an area of 50 x 50 pixels, therefore giving a total area of 55μm x 55μm, and a measurement time per map of around 2 hours. In some cases where more time was available, larger areas

were measured to acquire more data, followed by an extra pre-processing step to exclude any spectra measured outside the region of consensus pathology.

8.3. Data pre-processing

In-house software programs written in Matlab® R2014b (Mathworks, USA) were developed for pre-processing of Raman map data.

For larger maps run over large areas to maximise data acquisition, some of the measured areas were outside the region of consensus pathology. To ensure only the spectra within the region of interest were included, these maps were loaded into a software program that allowed visualisation of the map, and selection of pixels (i.e. spectra) for inclusion. An example of this is shown in Figures 8-2A and 8-2B.

Once the dataset included only spectra from regions with correctly identified tissue pathology, the wavenumber range is firstly truncated to 450-1800cm^{-1} to include only the fingerprint region of the Raman spectrum useful for biological samples. Spectral filtering is then performed to identify spectra that do not represent a true tissue measurement: this may be because of saturation of the detector due to high tissue fluorescence, low signal strength due to low laser power or some other system problem, or because a cosmic ray has been recorded at that pixel. This is done by calculating the mean and standard deviation (SD) of the dataset and applying thresholds based around these to exclude extreme outliers. For this analysis an intensity threshold of 5 SDs was used to identify areas of saturation (high intensity) or negligible signal (low

intensity), and gradient changes between adjacent wavenumbers of >15SDs to identify cosmic rays. These pixels were then removed from further analysis.

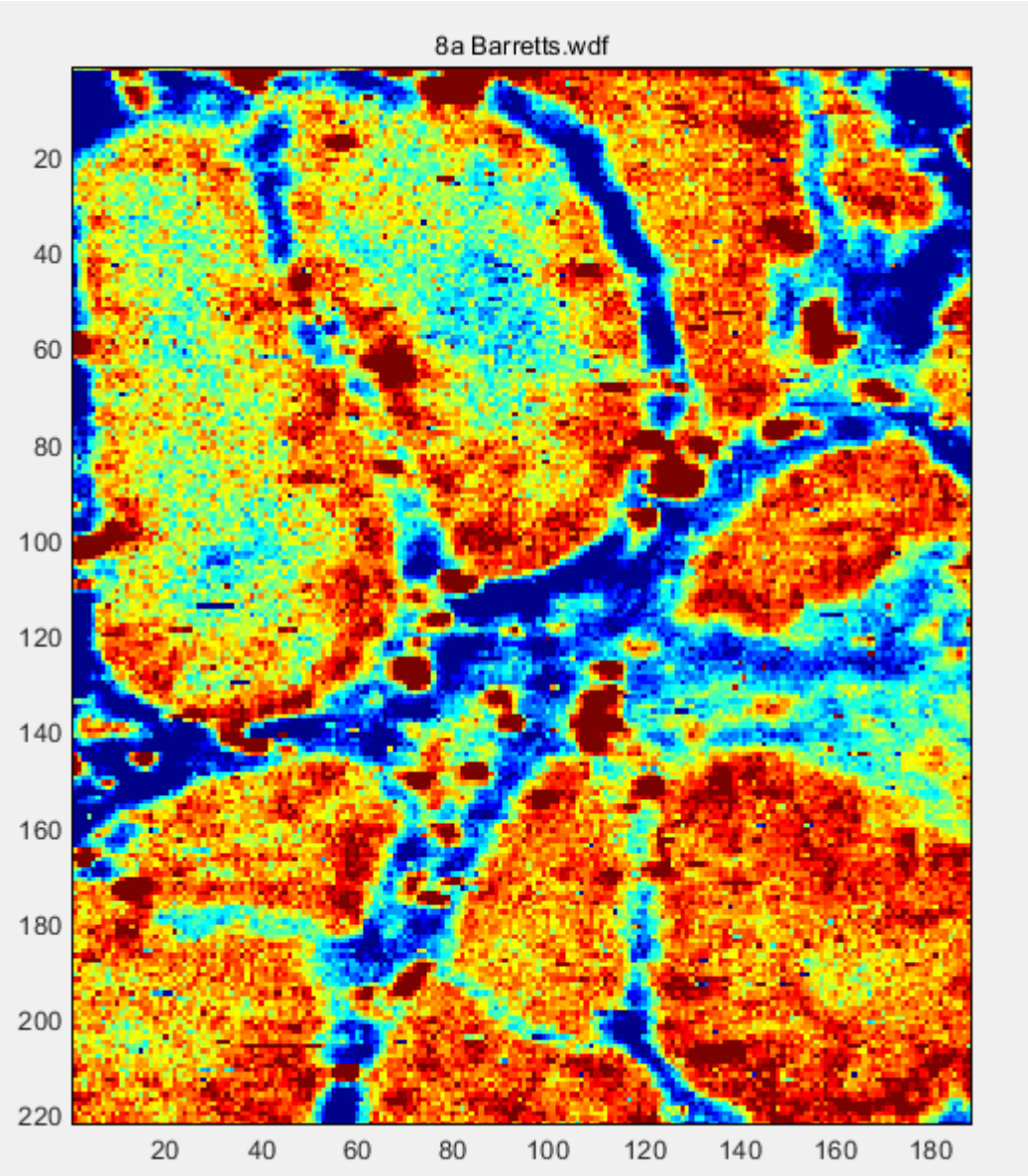


Figure 8-2A Selection of regions of interest within RaMP (in-house software) program. Adjusted intensity map to highlight glandular areas within a region of Barrett's tissue sample (see Figure 8-2B for region selection).

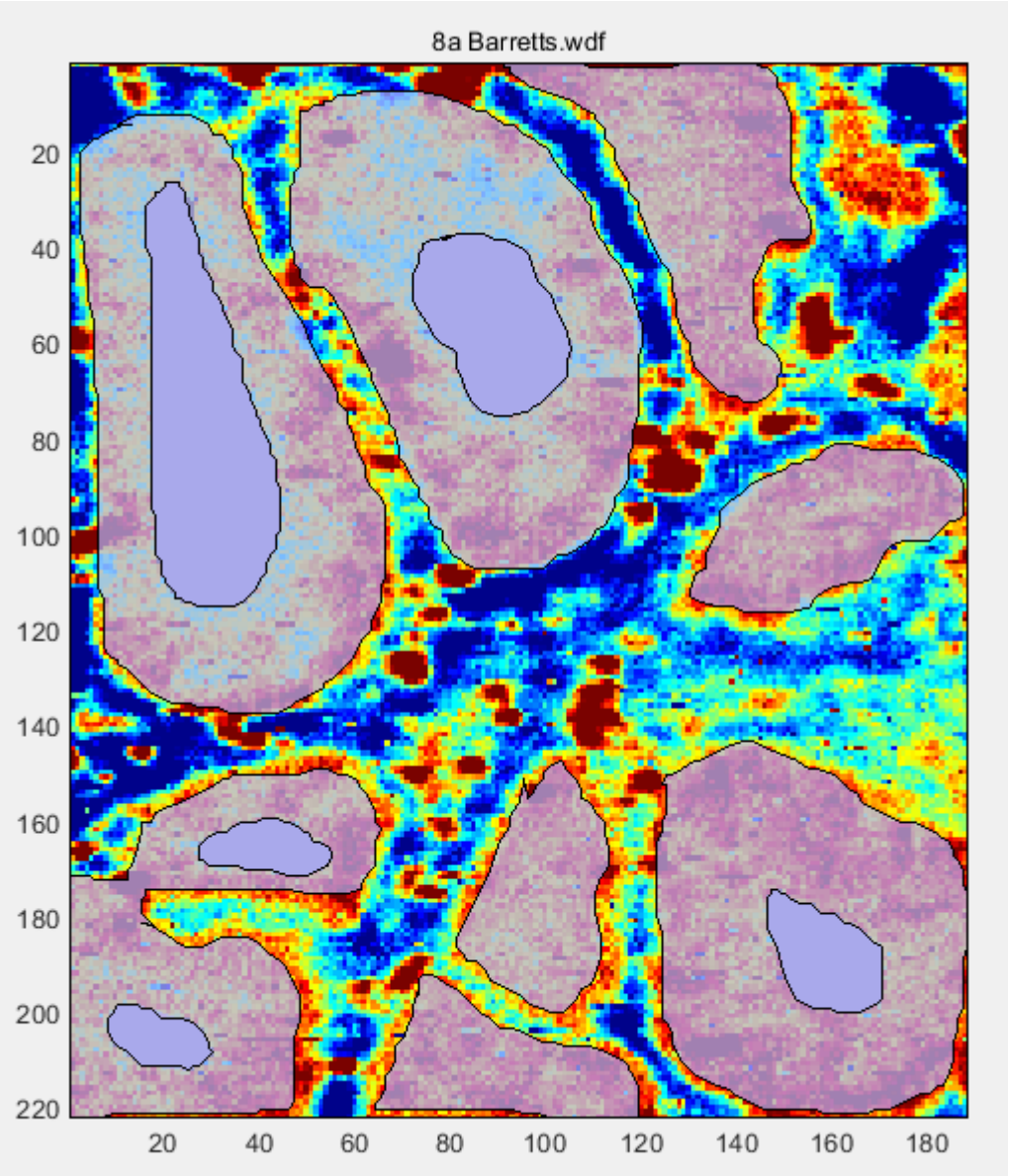
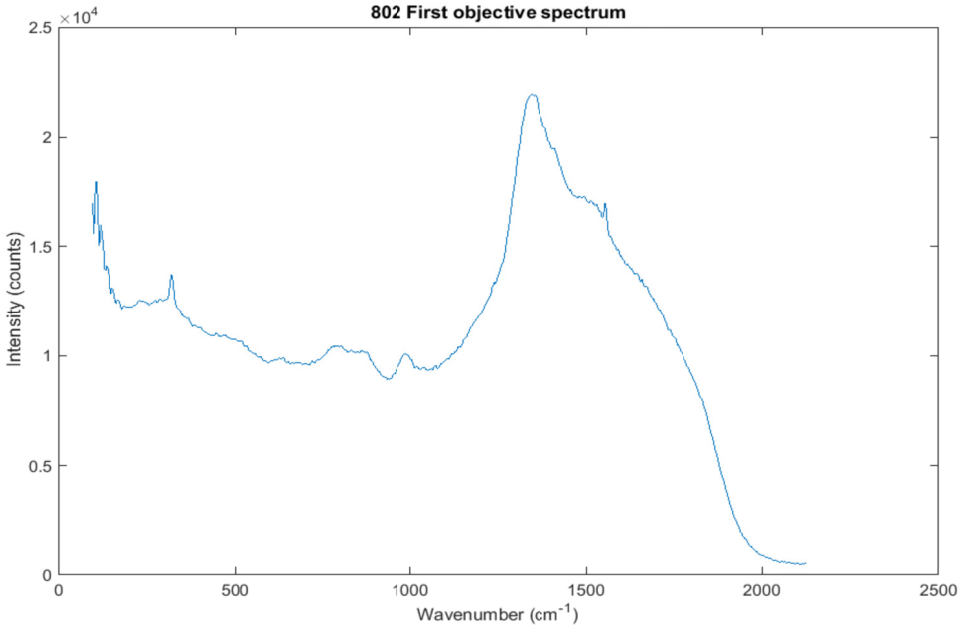


Figure 8-2B Selection of regions of interest within RaMP (in-house software) program. Grey areas show selected pixels to be included in data analysis (with darker grey areas corresponding to the gland lumen showing pixels not to be included). See Figure 8-2A for image without selected regions highlighted.

An EMSC correction algorithm was then applied which normalised the dataset to correct for variation in baseline signal intensity, removed the oxygen peak seen in all spectra at 1550-1570cm^{-1} (and interpolates to fill in the resulting gap in the spectrum) and corrects for background Raman signal generated within the objective lens.

A



B

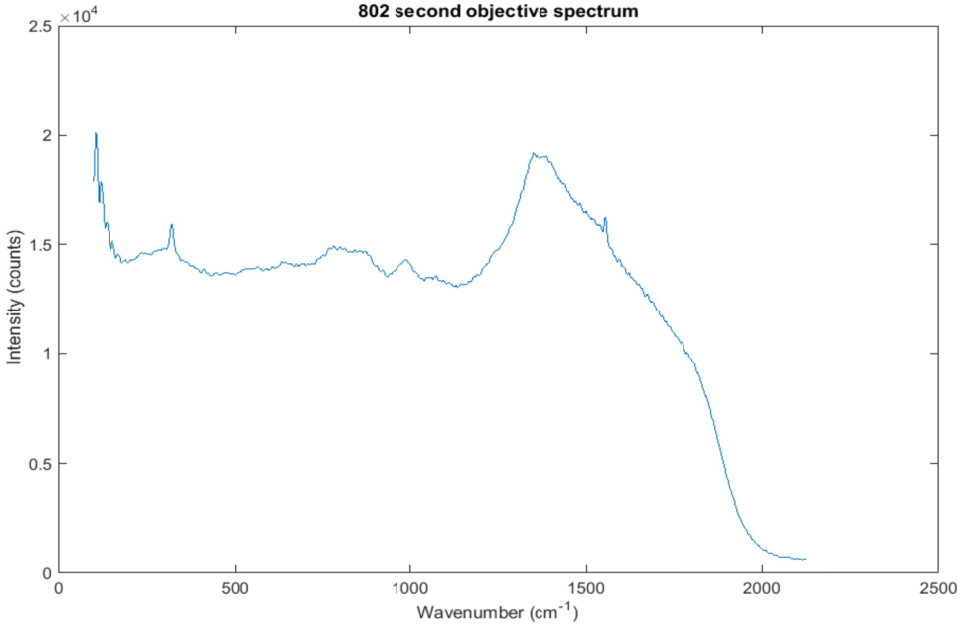


Figure 8-3 (A) Spectrum from first objective (used until April 2014) (B) Spectrum from second objective (used after April 2014).

Correcting for the objective was of particular importance in this project because there was a high background signal associated with the Leica x50 objective. As part of the development of the prototype instrument the objective lens was changed early on in the project because the first objective lens spectrum had a sharp peak in the key wavenumber region around 1400cm^{-1}. The second objective was, however, only a slight improvement. The spectra from the two objective lenses are shown in Figure 8-3 above.

This will have a large effect on the resulting spectrum measured from tissue, which will essentially be superimposed on the objective lens spectrum. The presence of a high Raman signal in the objective lens at a particular wavenumber may limit the ability to detect spectral features in a tissue signal at the same wavenumber. An objective lens that gives minimal Raman signal is therefore highly desirable.

Figure 8-4 shows a comparison of a number of objective lenses on another instrument: it is clear from these spectra that the Raman signal varies enormously with the objective lens.

In the examples in Figure 8-4, the x50 and x80 lenses have a very low Raman signal. The x5 lens has a small Raman signal, and the x2.5 lens has a large signal in the wavenumber region 600-1000cm^{-1}.

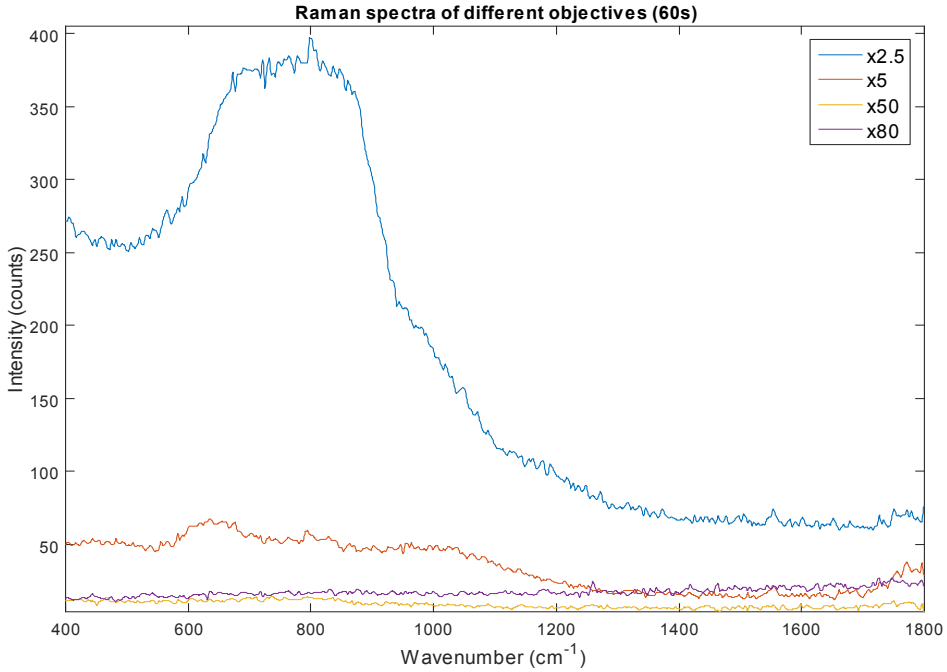


Figure 8-4 Raman spectra of a number of different objective lenses measured using same acquisition time on the same instrument. The Raman signal varies markedly between the lenses: this will affect the measured spectrum for any tissue sample. The x50 and x80 magnification show very little Raman signal.

8.4. Data analysis

The data analysis techniques used for FTIR mapping analysis (PCA, LDA and LOSOCV, described above in Section 4.7) were used here to develop classification models and explore differences between the pathology groups.

8.5. Summary of measurements

A total of 812,779 spectra were measured from 84 Raman maps, across 40 tissue samples, collected from 32 patients (some patients contributed tissue of

more than one pathology type, for example if an area of normal squamous tissue and an area of adenocarcinoma were collected from the same patient). The number of maps recorded and spectra measured for each pathology are shown in Table 8-1 below.

Table 8-1 Summary of 802 Raman mapping measurements

	No. of tissue samples	Total no. of Raman maps measured	Total no. of spectra measured	Total no. of spectra for analysis
Normal Squamous	10	23	310,545	186,088
Barrett's oesophagus	11	19	150,465	74,799
Dysplasia	10	23	139,255	101,322
Adenocarcinoma	9	19	212,514	170,323
TOTAL	40	84	812,779	532,532

After selecting regions of interest and excluding tissue areas that did not correspond to the correct pathology type, excluding cosmic rays, saturated spectra due to background fluorescence or spectra with low signal, a total of 280,247 spectra were excluded leaving 532,532 spectra for analysis. Mean spectra from each pathology group are shown in Figure 8-5.

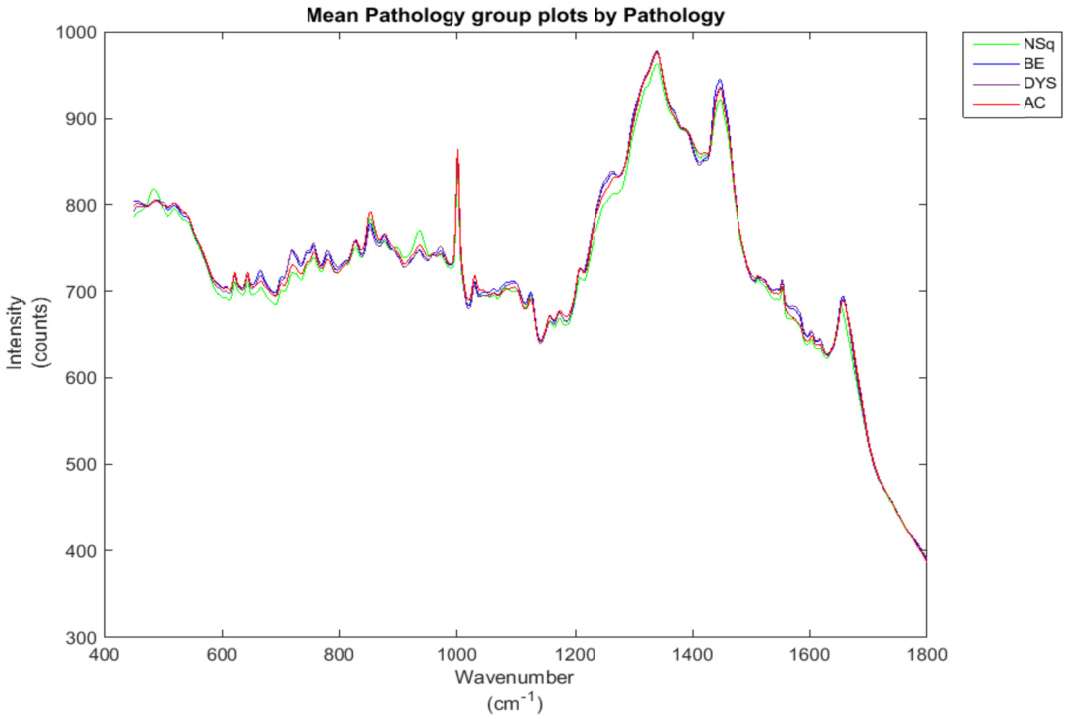


Figure 8-5 Mean spectra for each pathology group. NSq normal squamous; BE Barrett's oesophagus; DYS dysplasia; AC adenocarcinoma

Whilst some small differences between the mean spectra of the pathology groups can be seen, particularly for the normal squamous group, the most striking feature is that the mean spectra all show a very strong contribution from the objective lens, with a sharp, high peak around 1200-1400cm^{-1}. The differences between the spectra, along with the objective lens contribution, are discussed further below.

8.6. Classification models

In order to analyse the biochemical differences between all 4 pathology groups, the initial analysis plan was to build a 4 group classification model, then

examine spectral differences between the groups that contributed to classification. PCA-fed LDA was used to develop the model, which was then tested using LOSOCV. The performance of the 4 group classification model is shown in Table 8-2.

Table 8-2 Performance of the 4 group model: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology			
Predicted spectral pathology	**NSQ (no. of spectra)**	**BE (no. of spectra)**	**DYS (no. of spectra)**	**AC (no. of spectra)**
NSQ	**133,257**	2,613	4,975	6,890
BE	16,703	**28,556**	51,702	82,361
DYS	24,937	28,731	**27,433**	35602
AC	11,191	14,899	17,212	**45,470**
TOTALS	186,088	74,799	101,322	170,323

B

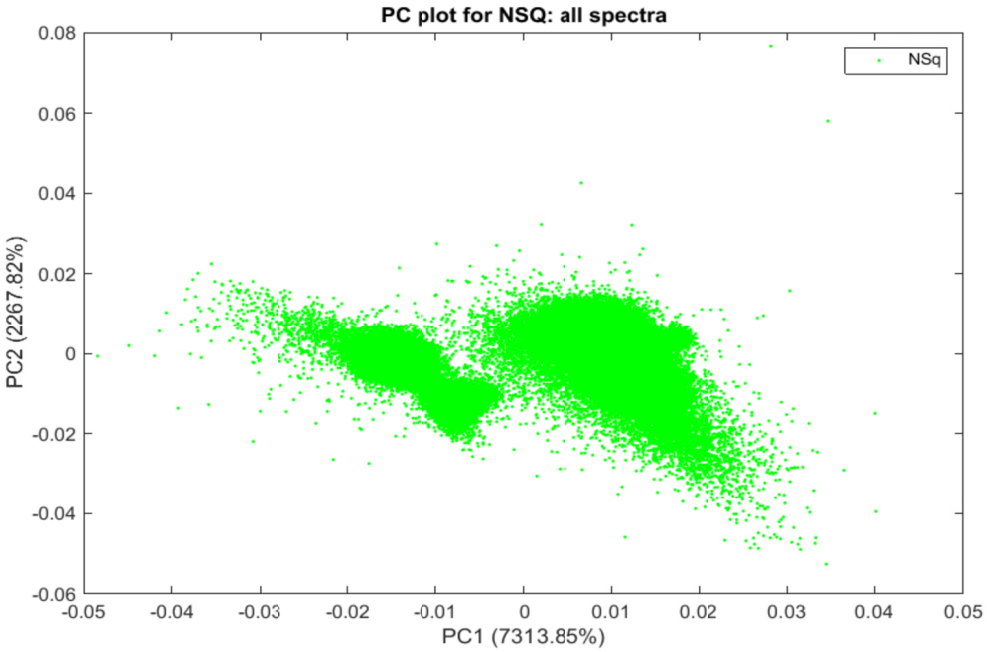
	Tissue pathology			
Performance of 4 group model	**NSQ**	**BE**	**DYS**	**AC**
Sensitivity (%)	71.6	38.2	27.1	26.7
Specificity (%)	95.8	67.1	79.3	88.0

The performance of the 4 group classification model was very poor. The model was able to classify normal squamous samples with reasonable accuracy, but classification of Barrett's, dysplasia and adenocarcinoma spectra was very poor. The dysplasia and adenocarcinoma sensitivity was only fractionally better than random assignment (which would be 25% in a 4 group model with equal numbers of spectra).

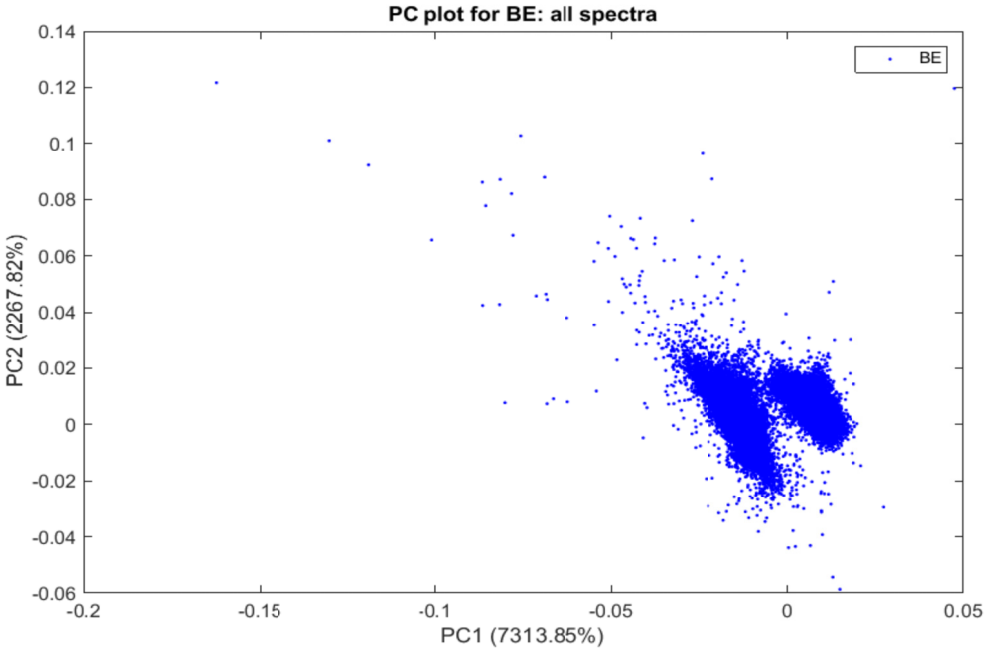
To analyse how the principal components (PC) were separating the data, the key PC plots for each pathology group were examined (Figure 8-6). Examination of the PC plots showed that each individual pathology separated into two groups on PC1 i.e. the principal component that described the greatest difference between spectra.

Given the separation of the pathologies into two groups, it seemed likely that this was caused by the use of two different objective lenses, and the failure to correct for this adequately with the objective correction algorithm. This would very seriously impair the classification ability. The dataset was then examined excluding those samples measured using the first objective. This resulted in the exclusion of 8 Raman maps (2 normal squamous, 4 Barrett's, 1 dysplasia and 1 adenocarcinoma) and 222,923 spectra (this relatively larger number of spectra given the number of samples reflects the fact that many of the earlier samples were very large maps, measured when there were fewer tissue samples and hence more time available for measurement). There were therefore 76 Raman maps and 309,609 spectra for analysis.

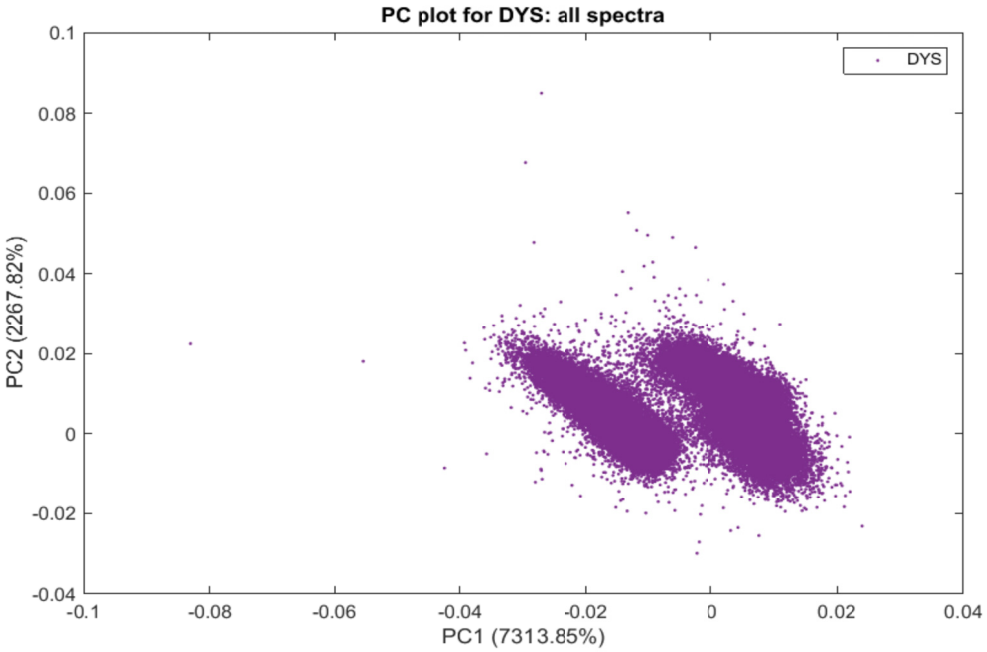
A


PC plot for NSQ: all spectra
NSq
PC2 (2267.82%)
PC1 (7313.85%)
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.05
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04
0.05

B


PC plot for BE: all spectra
BE
PC2 (2267.82%)
PC1 (7313.85%)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.2
-0.15
-0.1
-0.05
0
0.05

C



D

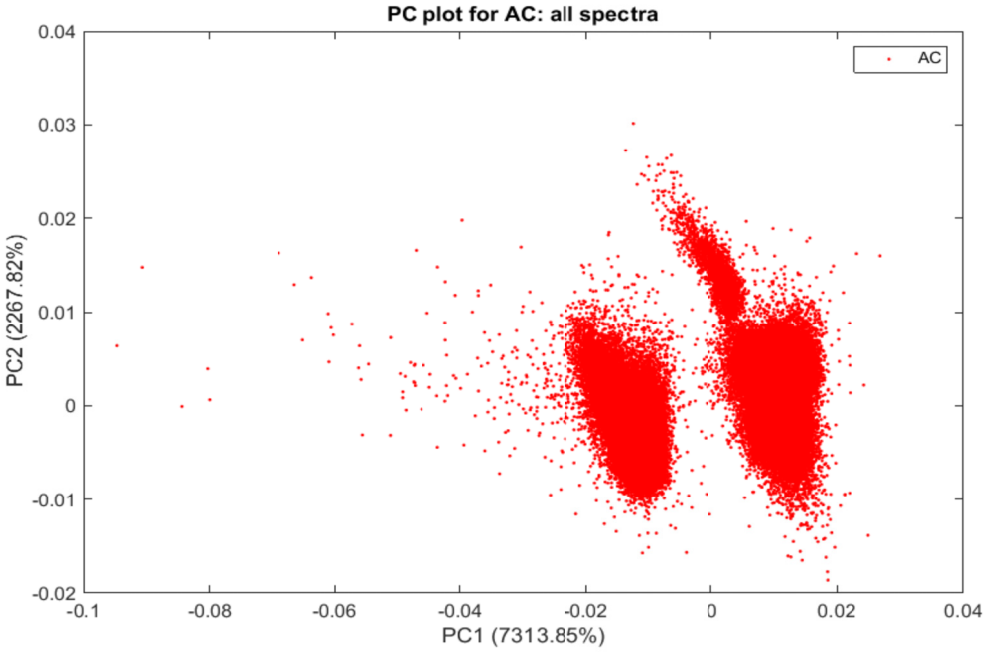


Figure 8-6A-D PC plots for PC1 and PC2 for each pathology. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

The classification model was re-run after exclusion of data measured on the first objective lens, and the performance is shown in Table 8-3, with the PC plots in Figure 8-7.

Table 8-3 Performance of the 4 group model after exclusion of first objective data: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

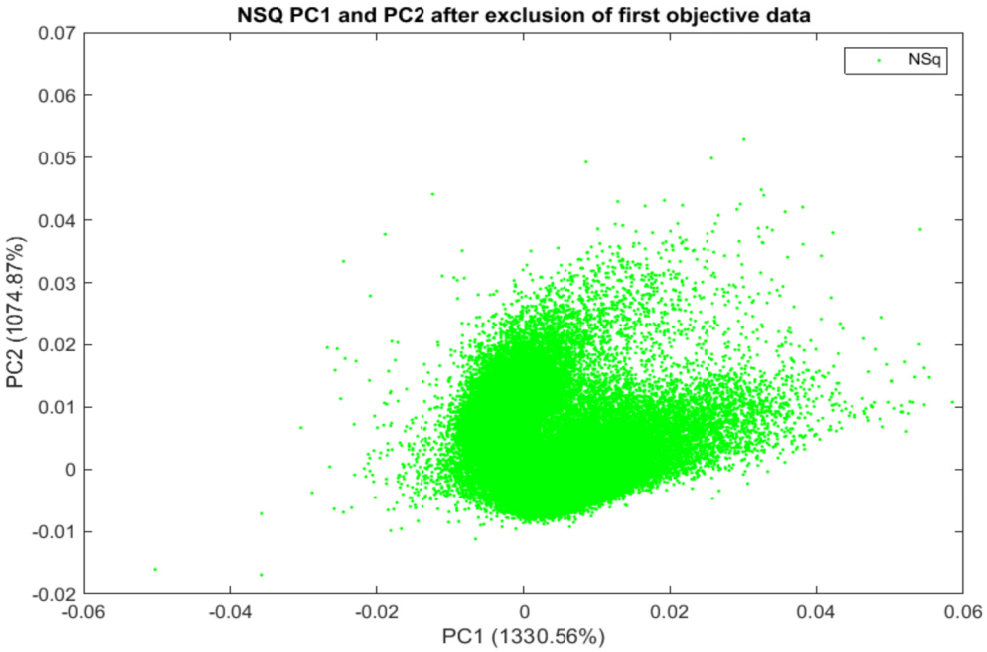
A

	Tissue pathology			
Predicted spectral pathology	**NSQ (no. of spectra)**	**BE (no. of spectra)**	**DYS (no. of spectra)**	**AC (no. of spectra)**
NSQ	**102,879**	1,046	3,307	3,247
BE	3,639	**20,620**	11,728	15,350
DYS	16,350	6,616	**27,017**	9,343
AC	8,365	7,826	10,510	**61,766**
TOTALS	131,233	36,108	52,562	89,706

B

	Tissue pathology			
Performance of 4 group model	**NSQ**	**BE**	**DYS**	**AC**
Sensitivity (%)	78.4	57.1	51.4	68.9
Specificity (%)	95.7	88.8	87.4	89.9

A



B

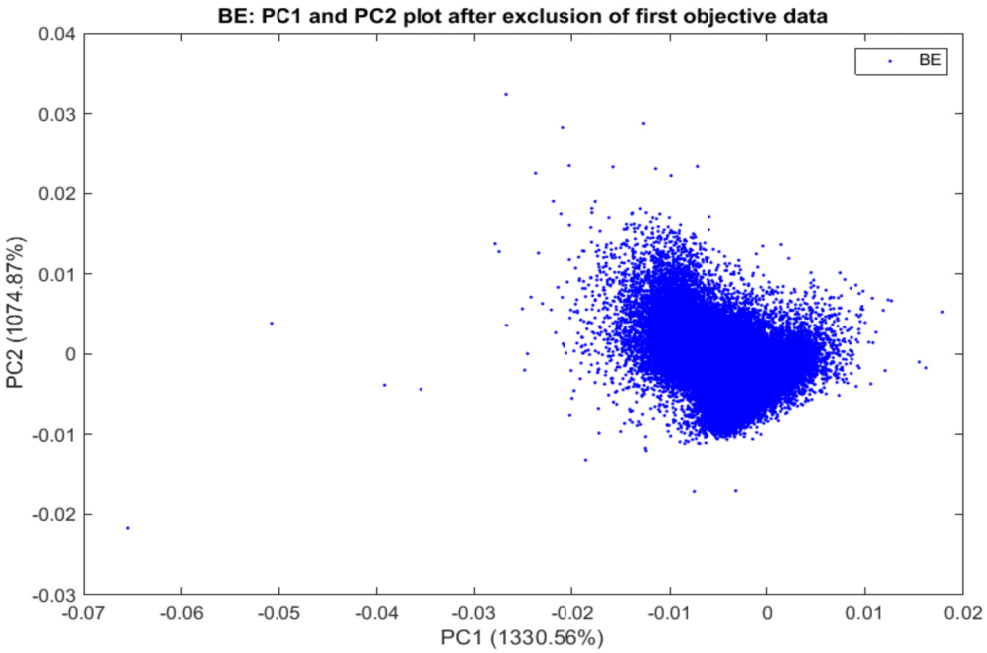
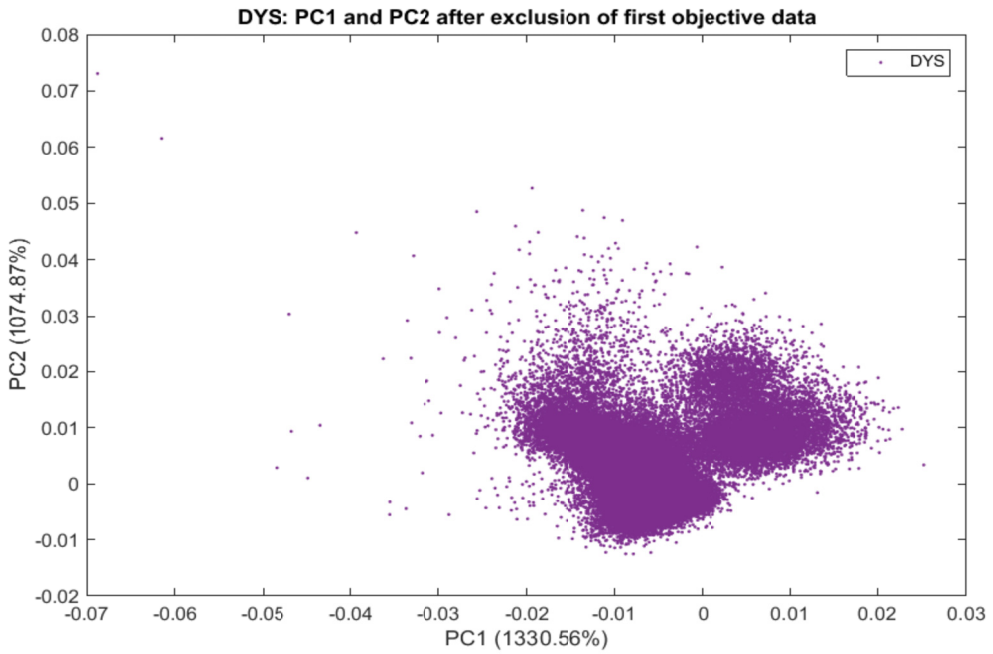


Figure 8-7A and B PC plots for PC1 and PC2 for each pathology after exclusion of the first objective data. A) Normal squamous (NSQ) B) Barrett's oesophagus (BE).

C



D

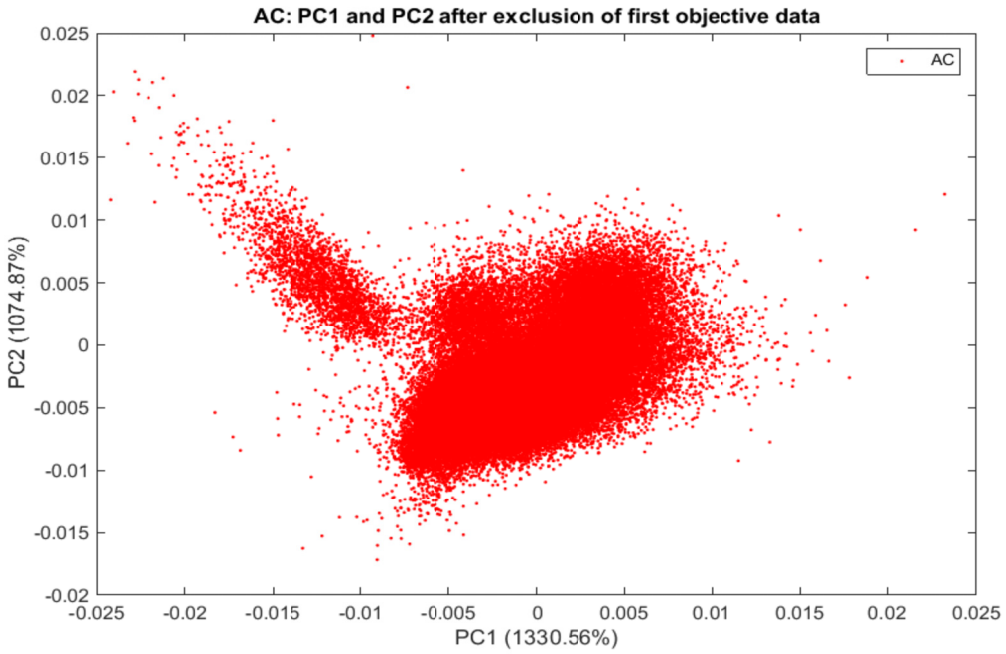


Figure 8-7C and D PC plots for PC1 and PC2 for each pathology after exclusion of the first objective data. C) Dysplasia (DYS) D) Adenocarcinoma (AC).

This showed a marked improvement but nonetheless the classification performance remained poor, suggesting that limited biochemical data would be learned from studying the differences between the groups.

The PCs in Figure 8-7 showed that there was no longer a clear separation of each pathology into two groups, but there was still considerable heterogeneity. The dysplasia and adenocarcinoma groups particularly demonstrated separation of some spectra from the main group. This suggested that certain sample measurements were systematically different from others within a pathology group, due to inaccuracies with either region selection or the instrument.

To exclude such differences, the maps were examined and excluded if the spectra from the map formed two separate groups on PC analysis, or appeared very heterogeneous, suggesting either two different tissue types due to inaccurate region selection that may not have corresponded to consensus pathology, or a change in the instrument during measurement. This decision process introduced an element of subjectivity, and 'over-exclusion' of outliers risks falsely elevating classification performance. However, this step was primarily to explore whether the performance could be improved sufficiently to justify further spectral analysis, given the poor performance of the classifier demonstrated thus far, rather than to test spectral classification.

This resulted in the exclusion of 9 further samples (3 normal squamous, 4 dysplasia, 2 adenocarcinoma), and 131,973 spectra, leaving a total of 177,636 spectra and 67 samples. Performance of the 4 group model after these further exclusions is shown in Table 8-4.

Table 8-4 Performance of the 4 group model after further exclusions: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology			
Predicted spectral pathology	**NSQ (no. of spectra)**	**BE (no. of spectra)**	**DYS (no. of spectra)**	**AC (no. of spectra)**
NSQ	**38,195**	1,039	215	1,106
BE	772	**18,923**	9,087	9,938
DYS	659	7,004	**24,875**	2,691
AC	4,325	9,142	8,786	**40,879**
TOTALS	43,951	36,108	42,963	54,614

B

	Tissue pathology			
Performance of 4 group model	**NSQ**	**BE**	**DYS**	**AC**
Sensitivity (%)	86.9	52.4	57.9	74.9
Specificity (%)	98.2	86.0	92.3	81.9

After the further exclusion process described above, the classification performance showed a minor improvement but remained poor overall. The model was able to classify the normal squamous spectra with good sensitivity and very high specificity, and the adenocarcinoma classification was also reasonably good.

Given the persistently poor performance of the 4 group model in discriminating between Barrett's, dysplasia and adenocarcinoma, a simpler analysis was

performed to see if the classifier was able to detect differences in a 2 group model: Barrett's versus the dysplasia/adenocarcinoma combined group. The results are shown in Table 8-5.

Table 8-5 Performance of the 2 group model after further exclusions: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology	
Predicted spectral pathology	**BE (no. of spectra)**	**DYS/AC (no. of spectra)**
BE	**22,878**	27,995
DYS/AC	13,230	**69,582**
TOTALS	36,108	97,577

B

	Tissue pathology	
Performance of 2 group model	**BE**	**DYS/AC**
Sensitivity (%)	63.4	71.3
Specificity (%)	71.3	63.4

The performance of the 2 group model remained poor. The strong background signal from the objective lens was a concern, so we aimed to test whether this was impairing the classification performance. The 2 group analysis was repeated, excluding those wavenumbers in the region most affected by the objective. The strongest peak from the objective lens was in the region 1250-1400cm^{-1}, and there was a strong effect in the wavenumber region above 1700cm^{-1}, so the truncated dataset excluded these wavenumbers.

Table 8-6 Performance of the 2 group model using truncated wavenumber range: (A) confusion matrix showing classification of individual spectra (with correctly classified spectra in **bold**), and (B) sensitivity and specificity of the model for classifying individual spectra. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

A

	Tissue pathology	
Predicted spectral pathology	**BE (no. of spectra)**	**DYS/AC (no. of spectra)**
BE	**22,788**	27,922
DYS/AC	13,320	**69,665**
TOTALS	36,108	97,577

B

	Tissue pathology	
Performance of 2 group model	**BE**	**DYS/AC**
Sensitivity (%)	63.1	71.4
Specificity (%)	71.4	63.1

This result was almost identical to the 2 group analysis using the full wavenumber range, implying that this wavenumber range from 1250-1400cm^{-1}, and above 1700cm^{-1}, did not contribute any spectral information to the classification model.

8.7. Spectral analysis

The 4 group models discriminated normal squamous spectra reasonably accurately, however the poor performance of the 2 group model suggested

there was limited useful spectral information with which to distinguish between the Barrett's and dysplasia/adenocarcinoma groups.

The result from the truncated dataset suggests that, in the region most affected by the objective signal, any spectral differences between tissue types are obscured.

It is therefore only possible to perform a limited spectral analysis, both in terms of wavenumber range, and the effect of the objective in obscuring small peaks throughout the spectrum due to shot noise (this is explained further in the Discussion below, section 8.8). Assigning biochemical constituents to spectral peaks is usually done tentatively: in the context of these limitations even greater caution must be advised in interpreting peaks in this dataset. Consequently we will focus only on the larger spectral differences in the subtraction spectra and avoid the wavenumber range which has been shown to contain no useful discriminatory information.

Looking at the differences in the subtraction spectra between normal squamous and the other pathology groups (Figure 8-8), the normal squamous group shows strongly positive peaks at 483 and 934cm^{-1} – these peaks are also sufficiently large that the difference in the normal squamous group can be seen clearly in the mean spectra in Figure 8-5 earlier in this chapter. Both of these peaks have been consistently reported in previous Raman studies in the oesophagus as representing glycogen, and present in greater amounts in normal squamous tissue (please refer to Table 8-7 below for further references of peak assignments reported in the oesophagus) (Almond et al. 2013; Hutchings et al. 2010; Shetty et al. 2006). There are a number of other glycogen peaks reported in the literature: glycogen peaks reported in multiple studies

include those at 852-5cm^{-1} and 1086-8cm^{-1}. Figure 8-8 shows positive peaks at 850-861cm^{-1}, and a small peak at 1081cm^{-1} – it is possible these correspond to glycogen. This finding of increased glycogen content in normal squamous tissue is therefore consistent with both the reported literature, and the findings from the FTIR mapping section in Chapter 4.

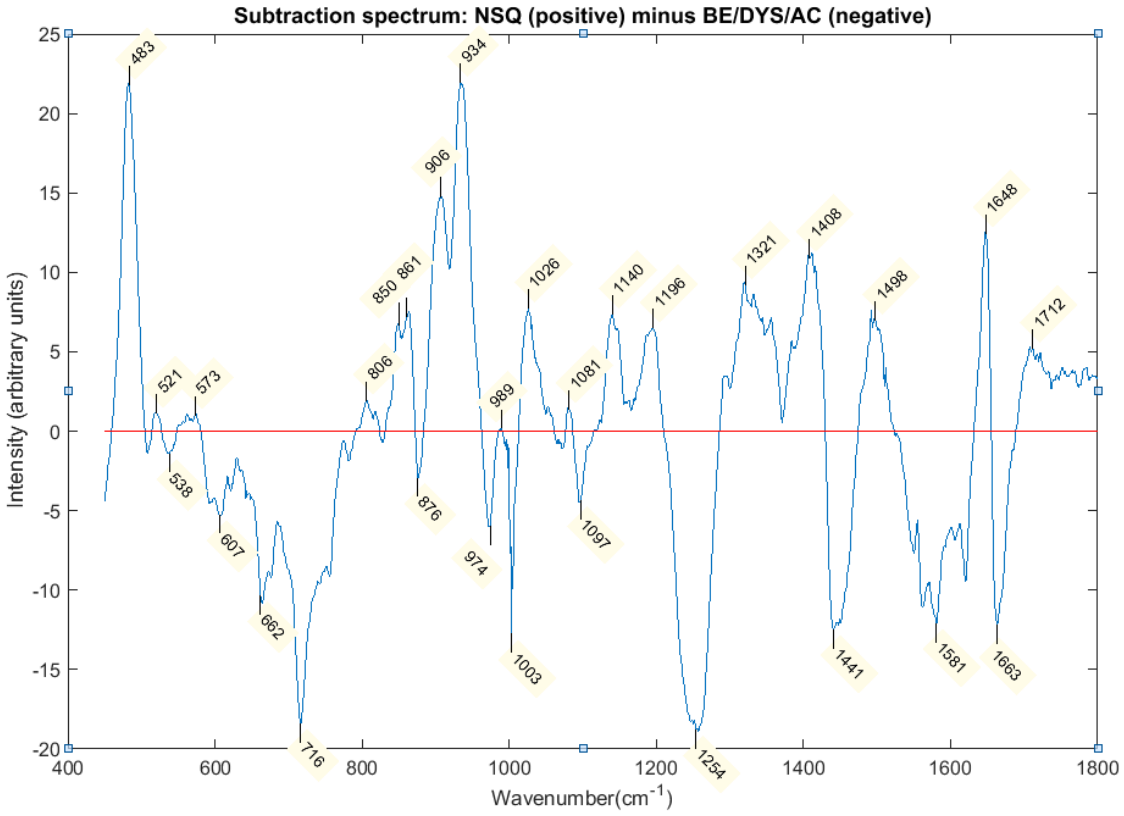


Figure 8-8 Subtraction spectra comparing normal squamous (NSQ) mean spectrum versus combined mean spectrum for Barrett's (BE), dysplasia (DYS) and adenocarcinoma groups.

The main amide I peak around 1650cm^{-1} splits to form two peaks at 1648cm^{-1} (positive) and 1663cm^{-1} (negative). This peak represents proteins, and therefore appears to be roughly equal across normal squamous and the other tissue types.

The major negative peaks represent higher content in the combined Barrett's/dysplasia/adenocarcinoma group, and occur at 716, 1003, 1254, 1441-50 and 1581cm^{-1}. DNA peaks have previously been reported at 719-20cm^{-1} and 1576-9cm^{-1} and it is possible that the increased DNA content of these tissues is reflected here. The 1003cm^{-1} (phenylalanine) and 1441-50cm^{-1} peaks are likely to represent increased protein content. The 1254cm^{-1} peak occurs at the beginning of the large peak in the objective spectrum and is likely to be artefact.

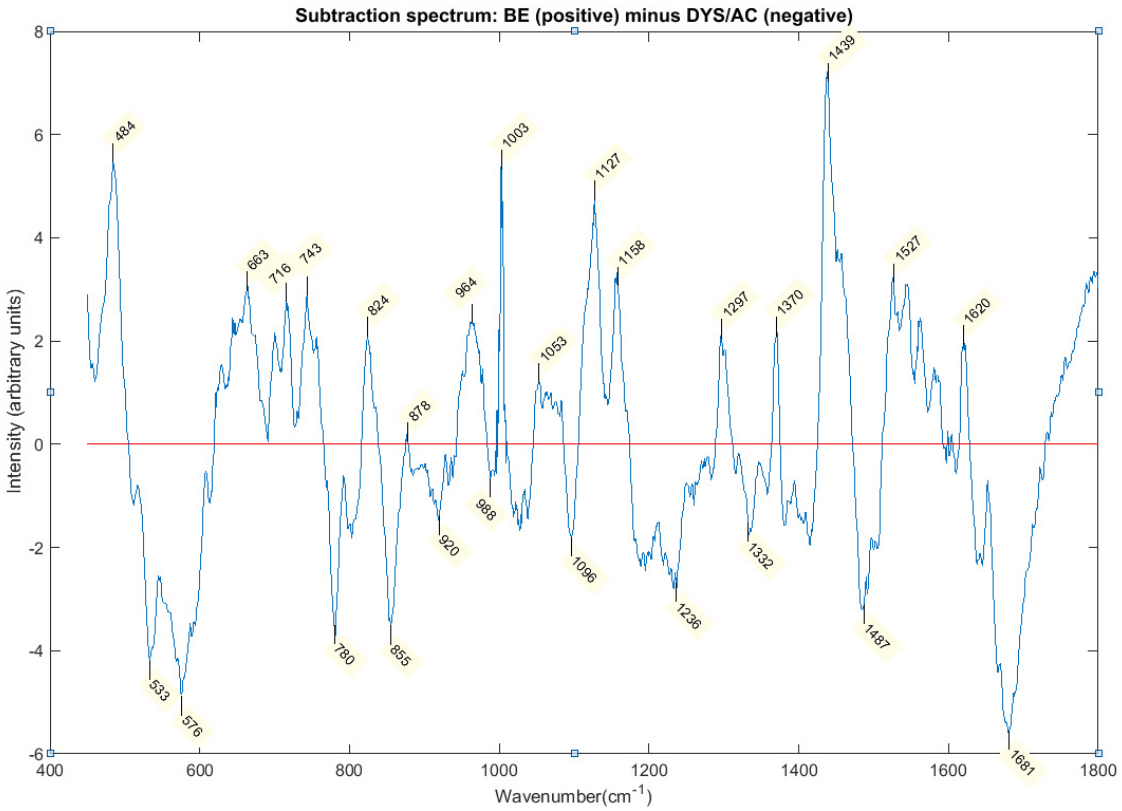


Figure 8-9 Subtraction spectra comparing Barrett's mean spectrum versus combined dysplasia and adenocarcinoma mean spectrum.

The differences in the subtraction spectra of the Barrett's versus the combined dysplasia/adenocarcinoma group (Figure 8-9) are small compared to those

seen in Figure 8-8 (the 'arbitrary units' in both figures are the same scale). Although the first major positive peak at 484cm^{-1} may represent glycogen, this is not borne out in the presence of other positive glycogen peaks (e.g. 855, 934, 1082cm^{-1}) elsewhere in the spectrum. Similarly, there does not appear to be a consistent pattern amongst the most commonly reported DNA peaks at 781-5, 852-5, 1334, 1576-9cm^{-1}. The positive peaks at 1003cm^{-1} and 1439cm^{-1} may reflect protein content, but the amide I peak around 1650cm^{-1} is predominantly negative, so again interpretation is difficult. In summary, this analysis is consistent with the inability of the classification model to identify useful spectral information with which to classify these pathology groups.

Table 8-7 Raman peak assignments in oesophageal tissue

Wavenumber (cm^{-1})	Biomolecule	Reference
470	Glycogen	Hutchings et al., 2010
481-90	Glycogen	Shetty et al., 2006, Almond et al., 2013
719-20	DNA	Shetty et al., 2006
748-55	DNA	Shetty et al., 2006
750	Lactic acid	Shetty et al., 2006
781-5	DNA	Shetty et al., 2006, Hutchings et al., 2010, Almond et al., 2013
820	Protein	Shetty et al., 2006
852-5	Glycogen Protein	Shetty et al., 2006, Hutchings et al., 2010 Hutchings et al., 2010
884	Protein	Shetty et al., 2006
885	DNA	Almond et al., 2013
918	Lactic acid	Shetty et al., 2006
933-7	Glycogen	Shetty et al., 2006, Almond et al., 2013
936-40	Proteins	Bergholt et al., 2014, Hutchings et al., 2010

944	Glycogen	Hutchings et al., 2010
1001-4	Phenylalanine	(Bergholt et al. 2014) (Stone et al. 2004) Hutchings et al., 2010
1018-21	Nucleic acids	Almond et al., 2013
1036	Glycogen	Hutchings et al., 2010
	Protein (Collagen)	Hutchings et al., 2010
1048	Glycogen	Shetty et al., 2006
1078	Lipids	Bergholt et al., 2014
1086-8	Glycogen	Hutchings et al., 2010, Almond et al., 2013
1128	Glycogen	Almond et al., 2013
1135	Glycogen	Hutchings et al., 2010
1223	Protein	Shetty et al., 2006
1261	Protein	Hutchings et al., 2010
1278	Protein	Shetty et al., 2006
1265	Protein (amide III)	Bergholt et al., 2014
1302	Lipid	Bergholt et al., 2014
1312	Protein	Hutchings et al., 2010
1333-8	Glycogen	Shetty et al., 2006, Hutchings et al., 2010
1334-5	DNA	Almond et al., 2013, Bergholt et al., 2014
1360	Nucleic acid	Almond et al., 2013
1453	Protein	Hutchings et al., 2010
1467	Glycogen	Hutchings et al., 2010
1511	Nucleic acid	Almond et al., 2013
1576-9	DNA	Bergholt et al., 2014, Hutchings et al., 2010
1618	Porphyrins	Bergholt et al., 2014
1655-9	Amide I (Protein)	Bergholt et al., 2014, Hutchings et al., 2010
1663	DNA	Shetty et al., 2006
	Protein	Shetty et al., 2006
1745	Lipids	Bergholt et al., 2014

8.8. Discussion

The aim of this study was to use high quality Raman mapping measurements with a long acquisition time to investigate spectral differences between 4 pathology groups in the oesophagus: normal squamous, Barrett's oesophagus, dysplasia and adenocarcinoma.

Discrimination of spectra from normal squamous tissue was performed with reasonable accuracy, and a sensitivity of 86.9% and specificity 98.2%. The largest spectral differences contributing to this classification were strong glycogen peaks at 483 and 936cm^{-1}, consistent with previously reported work in the oesophagus, and the findings from the FTIR mapping study in Chapter 4 (Almond et al. 2013; Bergholt et al. 2014; Hutchings et al. 2010; Shetty et al. 2006). Tentative peak assignment suggested increased DNA and protein content in the Barrett's/dysplasia/adenocarcinoma tissue compared to normal squamous, and this again is consistent with previous findings.

The discrimination of the Barrett's, dysplasia and adenocarcinoma groups was very poor. The paucity of useful spectral differences for classification was confirmed by the absence of consistent biochemical differences on subtraction analysis.

The diagnostic classification performance in this study was worse than reported previously for Raman in the oesophagus, both in earlier work by our group and that reported by others (Almond et al. 2013; Bergholt et al. 2014; Hutchings et al. 2010; Kendall et al. 2003). Since this result was rather worse than previous studies using similar, or shorter, acquisition times, we had to consider what had changed within the new prototype system that could cause such a poor result. The most obvious difference was the Leica objective lens, which had a very

strong background Raman signal. The previous studies had used either an Olympus objective lens or a probe with a custom made lens to focus the laser beam.

The truncated 2 group model results imply that the objective lens is seriously limiting the classification performance in the wavenumber range where the largest peak is seen, between 1250-1400cm^{-1}, with this region contributing effectively no spectral information to the classification. It is likely that this region contains some key discriminatory spectral peaks: earlier work in the oesophagus has reported a protein peak at 1265cm^{-1} (Bergholt et al. 2014), a lipid peak at 1302cm^{-1} (Bergholt et al. 2014), a further glycogen peak at 1333-8cm^{-1} (Hutchings et al. 2010; Shetty et al. 2006), nucleic acid peaks at 1335cm^{-1} (Bergholt et al. 2014) and 1360cm^{-1} (Almond et al. 2013).

However, whilst the loss of spectral information in this region will impair classification, there are many peaks in other wavenumber regions that could potentially discriminate (as discussed in the spectral analysis section). We must therefore look for other reasons why our model failed to achieve good classification between Barrett's and dysplasia/adenocarcinoma tissue.

Although the largest peak in the objective lens spectrum is at 1250-1400cm^{-1}, there is high signal throughout the fingerprint range, with a number of other small peaks. Inaccuracies in the algorithm correcting for the objective may reduce the effect of spectral differences as a result.

A further problem associated with a high background Raman signal due to the objective lens is an increase in noise. Even with perfect correction for the objective lens signal, with a high signal you have an increase in shot noise which impairs the signal to noise ratio. Shot noise results from the quantum

nature of Raman scattering, and the random probability of scattering occurring for a given incident photon at a certain wavelength. The shot noise varies with the square root of the total number of counts, so as the total count increases, shot noise increases, but it becomes a smaller fraction of the signal.

Since the shot noise is proportional to the square root of the total number of counts, with a large background signal from the objective, this gives a large shot noise relative to the size of the measured tissue signal. Therefore even correcting for the objective lens still leaves a spectrum with a relatively high shot noise. The random variation in spectra due to noise can therefore overwhelm the true variation between the groups and seriously impair classification performance. As a result of this work other members of the Biophotonics group have developed a simulated model to demonstrate the effects of shot noise with a strong lens objective signal. This simulation is included in Appendix X.

There are a number of other factors that may have contributed to the poor performance of the classification model in this study. In this study, single pathologist reporting was used to classify pathology. Subsequent review by a second pathologist showed a number of discrepancies: the overall agreement for samples included in this study was 81.7% (see Appendix VII). If differences between LGD and HGD are ignored (since for this study we amalgamated these groups) and it is also assumed that ‘malignant: type unsure’ represents adenocarcinoma, then the agreement rises to 90%. The inclusion of non-corcordant samples may be a significant source of inaccuracy. If a reliable classification model could be constructed without these samples, it would be interesting to then apply it to these samples and then see how our model classifies them.

The failure of the correction algorithm to account for the change in objective lens resulted in the exclusion of a significant proportion of the total samples and measured spectra.

Further exclusions attempting to remove data that included multiple pathology groups or instrument problems led to a further reduction in the size of the dataset.

This was primarily stimulated by a concern about the functioning of the prototype Raman microspectrometer. On a number of occasions there had been repeated problems with the automated calibration procedure. No maps were run when the calibration failed, however given the problems with calibration we were concerned that the laser could have been running at low power or with suboptimal system alignment when the automated process reported the calibration had ‘passed’. There were a number of occasions where the system crashed during measurement and the map had to be repeated. Whilst none of the maps in which there was a known system failure were included, it is possible that a more subtle change in system alignment could have occurred without it being obvious in the resulting map, and the map would have been included.

These problems are related to the fact that this was a prototype system, and development is ongoing. We have worked closely with the manufacturers during this project to feed back the problems we have encountered, many of which have now been improved. The choice of objective lens has also been discussed at some length, and on the basis of this work the manufacturers plan to change the lens for future prototypes.

Other factors which could have limited the performance of the model could have been related to the sample biochemistry or to mislabelling of pathology group for spectra within the training model. The inclusion of the same patient in multiple pathology groups (e.g. where normal squamous or Barrett's samples were taken from those regions in a patient who also (at a separate location) had an adenocarcinoma), could present difficulty. If patient differences (as opposed to pathology differences) were being used by the classifier to discriminate, this would have impaired the performance of the model.

Alternatively, if there was genuine biochemical field change in these regions (i.e. the tissue adjacent to the tumour appeared histologically as Barrett's but contained genetic and biochemical changes consistent with adenocarcinoma), this would have led to misclassification.

As with the FTIR classification, LOSOCV was used to test the model, but multiple samples from the same patient were in the training set. Using a leave one patient out cross validation may further degrade the classification model if between-patient differences were being used as a basis of classification. This was not pursued further here since the primary aim was not to test classification ability, and the classification performance had been relatively poor anyway.

Future work will be discussed in the summative discussion in Chapter 10.

CHAPTER 9: COMPARISON OF RAMAN MAPPING AT 785NM, 830NM AND FTIR MAPPING

9. Comparison of Raman mapping at 785nm, 830nm and FTIR mapping

9.1. Study design

This project aimed to compare classification performance of spectroscopic mapping using different spectroscopic modalities (infrared and Raman) and different excitation wavelengths for Raman (785nm and 830nm) on the same tissue section. As a related project, contiguous tissue sections to those measured here were taken for measurement using CARS, to allow a further arm of comparison with another Raman modality, as part of a project by another PhD student at Exeter (Kelly Curtis).

The FTIR mapping data are presented in Chapter 4 above. In order to allow comparison, immediately after FTIR measurement (<24 hours) the same tissue region was then measured using Raman, first at 830nm then at 785nm. The FTIR methods and analysis are presented above, and the Raman methods and analysis are detailed below, with the FTIR data used here for comparative purposes. The CARS results are not included in this thesis.

9.2. Instrumentation

Raman maps were measured using a Renishaw InVia micro-spectrometer in the Department of Physics at Exeter University. This instrument is an older version of the 802 system described in Chapter 8 above. The InVia spectrometer uses the same Streamline^{TM} technique to raster scan across samples and create a Raman map (see Figure 8-1 in Chapter 8).

The InVia system has an in-built Neon-Argon lamp and uses this to optimise system alignment as described for the 802. Other calibration steps are not automated, and performed by the user each day prior to measurement.

Silicon calibration is performed manually, with measurement of a silicon standard followed by manual entry of the required wavenumber offset.

A measurement of a standard sample of green glass was then performed. Green glass has a smooth fluorescent spectral signal with few Raman peaks across a broad wavenumber range. This can be used to correct for variations in spectrometer detector sensitivity by applying an energy transfer function correction.

Cyclohexane was also measured as it has Raman peaks at 801, 1027, 1264 and 1441cm^{-1}, and drift in these peaks over time (e.g. due to changes in laser wavelength) can be monitored.

9.3. Sample measurement

The samples were first measured on FTIR, then Raman at 830nm then 785nm, with the order of measurement chosen to minimise the risk of thermal tissue damage with sequential measurements.

For Raman measurement the region selection had already been performed for FTIR. The white light image taken on the FTIR microscope was used to identify this area once the sample was mounted on the Raman system, and allow measurement of the same region of tissue with Raman.

The regions measured using FTIR were 128 x 128 pixels using 1.1μm pixel size, i.e. 140.8μm x 140.8μm. Since the closest step size on the InVia Raman system was 1.4μm, the maps were collected over 101 x 101 pixels, giving a total area of 141.4μm x 141.4μm. This is a very small difference in size between the maps, and likely to be smaller than the margin of error in identifying and measuring the exact same region of tissue on both systems.

Measurements at both excitation wavelengths were performed using a grating of 300l/m, with collected wavenumbers centred at 2350cm^{-1}, giving a wavenumber range 406.25-3738.55cm^{-1}. Inclusion of the high wavenumber range was selected because this would be helpful in informing the CARS measurements on the contiguous tissue sections as part of a separate PhD project.

An acquisition time of 30s per spectrum was used. For a laser line measuring 70μm, using a step size of 1.4μm and therefore 50 pixels, this gives an acquisition time 30/50 = 0.6s per pixel. This was chosen because it roughly equates to the target acquisition time for the Raman probe, of around 0.5s (previous work has shown good discrimination at 1s for an earlier prototype). It was also felt to be the shortest realistic time to get quality discrimination, and therefore comparable on that basis to the FTIR maps, though the time taken per map (12 minutes for FTIR, around 2 hours for a Raman map) were markedly different.

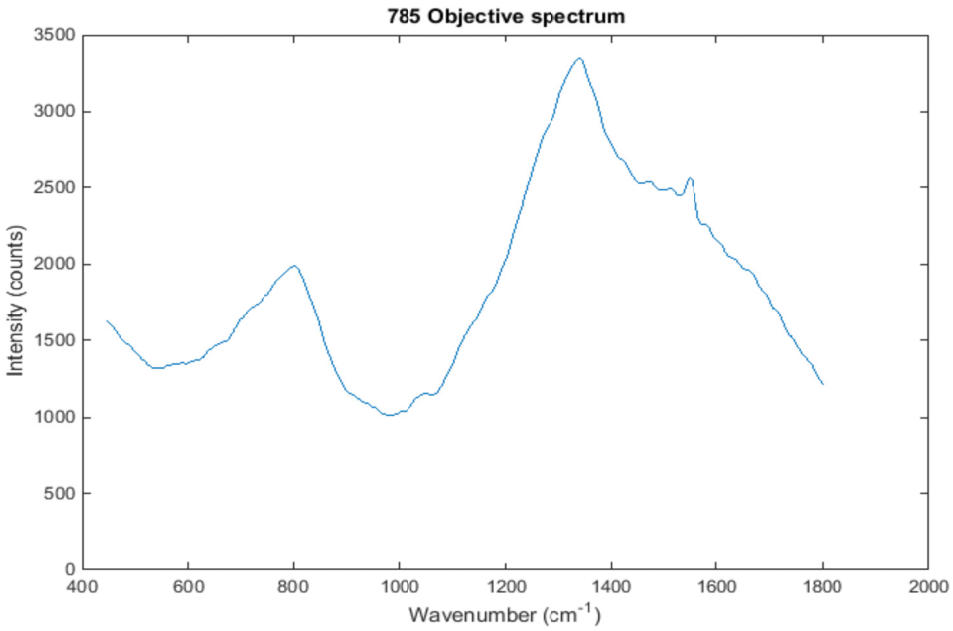
9.4. Data pre-processing

An in-house software programme written in Matlab® R2014b (Mathworks, USA) was used to identify cosmic rays and saturated spectra. Each Raman map was loaded into the software programme and the wavenumber range was truncated to 450-1800cm^{-1}, to leave only the 'fingerprint' region. Cosmic rays were removed by applying a median filter – this analyses a region of 3x3 pixels and replaces the central pixel with the median spectrum of the group, thus ensuring each pixel is unaffected by extreme outliers such as a cosmic ray. The map was inspected to look for large regions of saturation where the median filter would not work in removing outliers, and such areas were manually selected for removal.

The spectra were then normalised and an objective correction applied. The objective correction was again important as the Leica objective lens on the InVia system also has a strong contribution to the signal. The objective lens spectra measured at 785nm and 830nm are shown in Figure 9-1.

Measured at 785nm, the objective lens spectrum has a significant peak at 700-900cm^{-1}, and a very large, sharp peak from 1100-1500cm^{-1}. Measured using 830nm, the objective lens has a high intensity signal at wavenumbers below 1000cm^{-1}, with a very large peak from 700-900cm^{-1}, then the signal rapidly falls to a much lower level above 1200cm^{-1}.

A



B

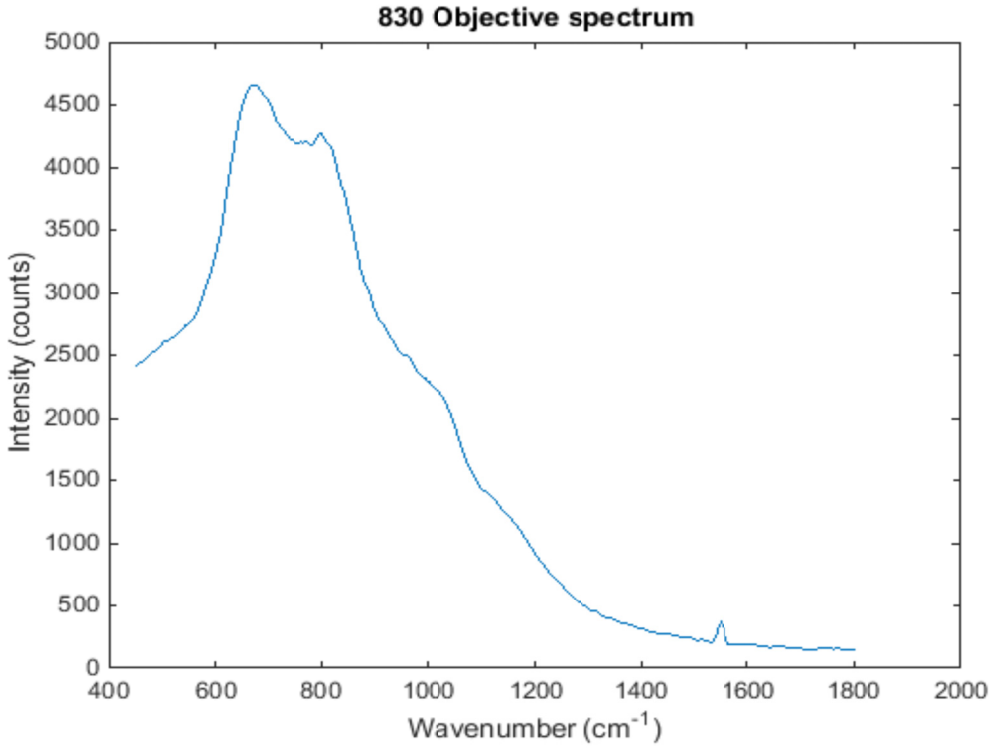


Figure 9-1 Objective spectrum measured at (A) 785nm and (B) 830nm.

9.5. Data analysis

The same data analysis techniques used for FTIR mapping analysis (PCA, LDA and LOSOCV, described above in Section 4.7) were employed for the Raman data to develop classification models and explore differences between the pathology groups. As with the FTIR, 50 PCs were calculated, with ANOVA threshold 0.99 used for inclusion in the training model.

9.6. Summary of measurements

Each of the tissue sample regions measured using FTIR was measured using 785nm and 830nm, i.e. 22 tissue samples from 19 patients (some patients contributed tissue of more than one pathology type, for example if an area of normal squamous tissue and an area of adenocarcinoma were collected from the same patient). 141.4μm x 141.4μm regions were measured to match the FTIR regions, with 2 maps measured per sample (for the normal squamous tissue samples one 'extra' map was measured for 3 samples, and one map crashed during measurement). At 785nm a total of 459,045 spectra were measured from 45 maps (one map from one of the dysplasia samples crashed during measurement); at 830nm a total of 469,246 spectra were measured from 46 maps. The number of maps recorded and spectra measured for each pathology, at 785nm (Table 9-1) and 830nm (Table 9-2) are shown below.

Table 9-1 Summary of InVia Raman mapping measurements at 785nm

	No. of tissue samples	No. of Raman maps	No. of spectra	No. of spectra after exclusions
Normal Squamous	5	12	122,412	118,775
Barrett's oesophagus	6	12	122,412	118,173
Dysplasia	5	9	91,809	88,944
Adenoca.	6	12	122,412	114,590
TOTAL	22	45	459,045	440,482

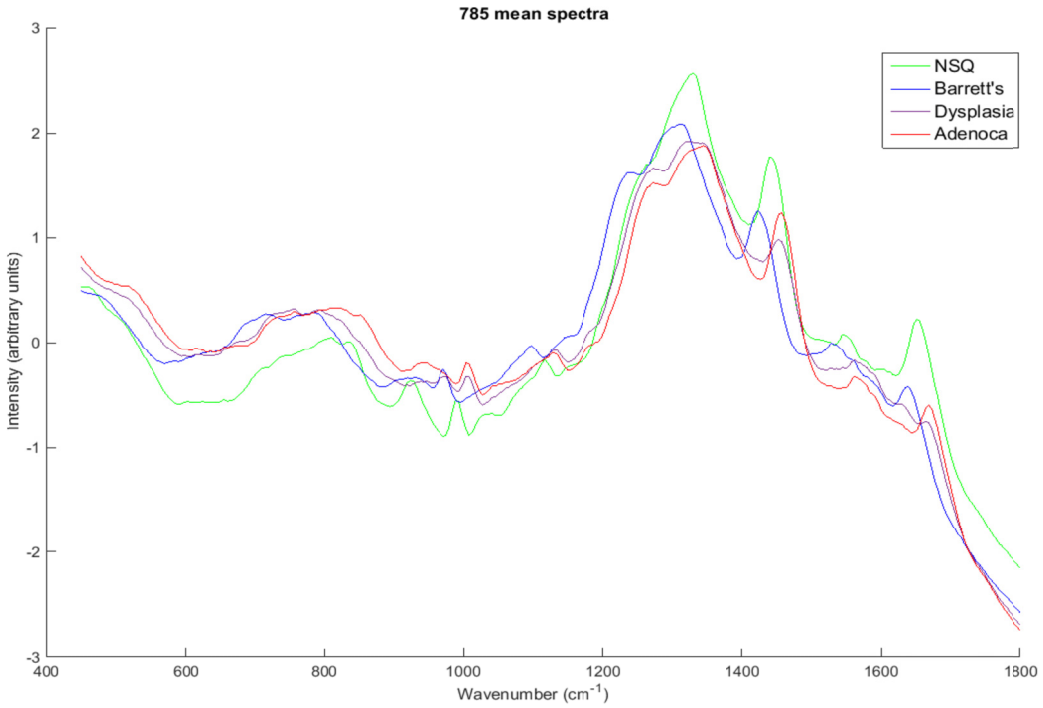
Table 9-2 Summary of InVia Raman mapping measurements at 830nm

	No. of tissue samples	No. of Raman maps	No. of spectra	No. of spectra after exclusions
Normal Squamous	5	12	122,412	118,943
Barrett's oesophagus	6	12	122,412	117,743
Dysplasia	5	10	102,010	98,521
Adenoca.	6	12	122,412	114,339
TOTAL	22	46	469,246	449,546

9.7. Spectral analysis and classification models

The mean spectrum for each pathology group measured at 785nm and 830nm is shown in Figure 9-2. The mean spectra have been normalised but not corrected for baseline or objective signal.

A



B

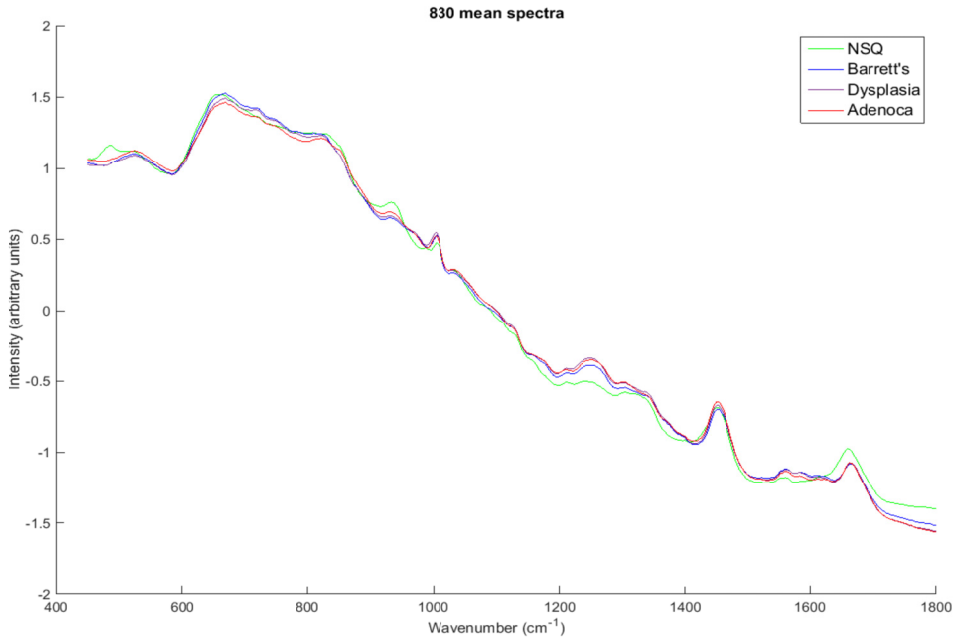


Figure 9-2 Mean spectrum for each pathology group measured at (A) 785nm and (B) 830nm. The spectra are normalised, but have not been objective or baseline corrected.

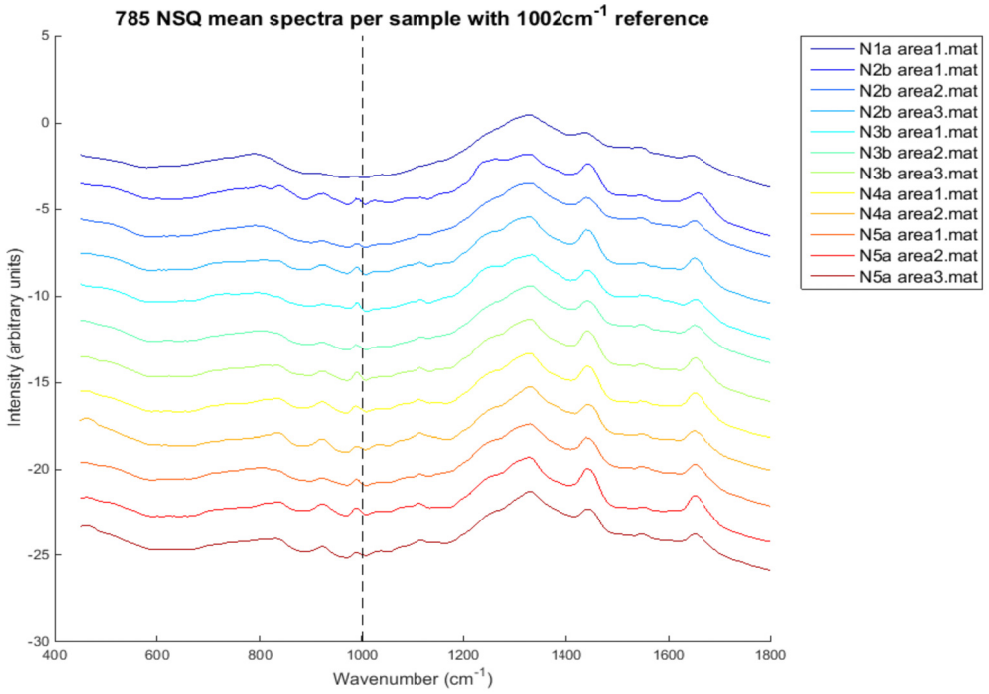
The 785nm mean spectra show strong peaks from 700-800cm^{-1} and 1200-1400cm^{-1} as seen in the objective spectrum above. However it appears the spectral peaks are shifted relative to one another, with a similar spectral shape but major peaks occurring at markedly different wavenumbers.

The 830nm mean spectra show a very strong effect from the objective – with a high peak around 600-800cm^{-1}, then a marked drop in the baseline intensity as the wavenumber increases. There also appears to be a more subtle peak shift with respect to the normal squamous samples.

The appearances of the mean spectra were concerning: firstly, it suggested that the objective lens contribution may again be a problem as the mean spectra at both wavenumbers exhibited a strong effect from the objective spectrum, and secondly, the presence of a peak shift suggested a calibration problem, with potential to be a major source of variation in the data, which would affect classification performance.

To identify which of the individual samples showed evidence of peak shift, the sample mean plots for each pathology group were examined individually (shown in Figures 9-3 to 9-6). The 1002cm^{-1} peak represents phenylalanine and is likely to be relatively constant across the pathology groups (Dingari et al. 2012; Stone et al. 2004). This peak can be used as a reference standard within each sample, and the position of the peak used to identify peak shift if it deviates from 1002cm^{-1}.

A



B

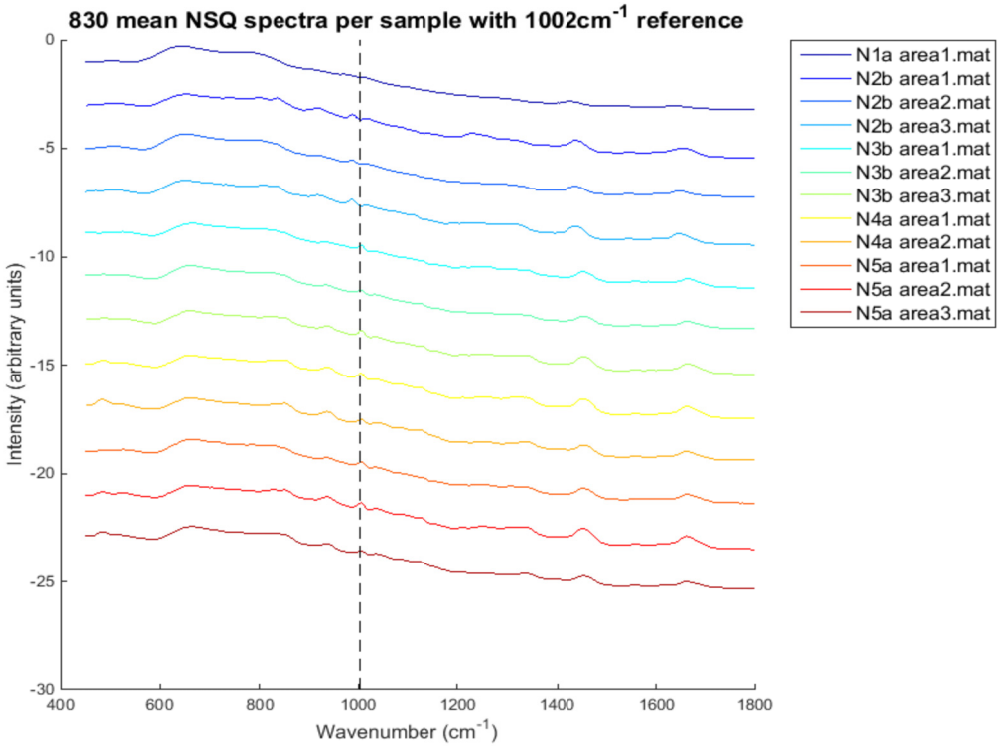
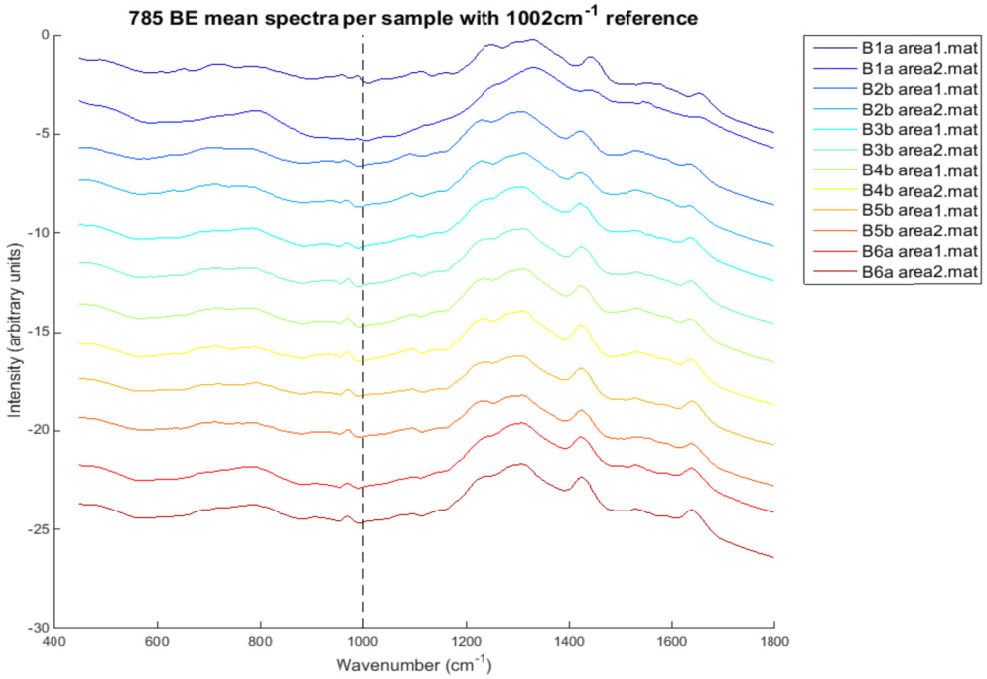


Figure 9-3 Sample mean spectra for normal squamous (NSQ) samples at (A) 785nm and (B) 830nm.

A



B

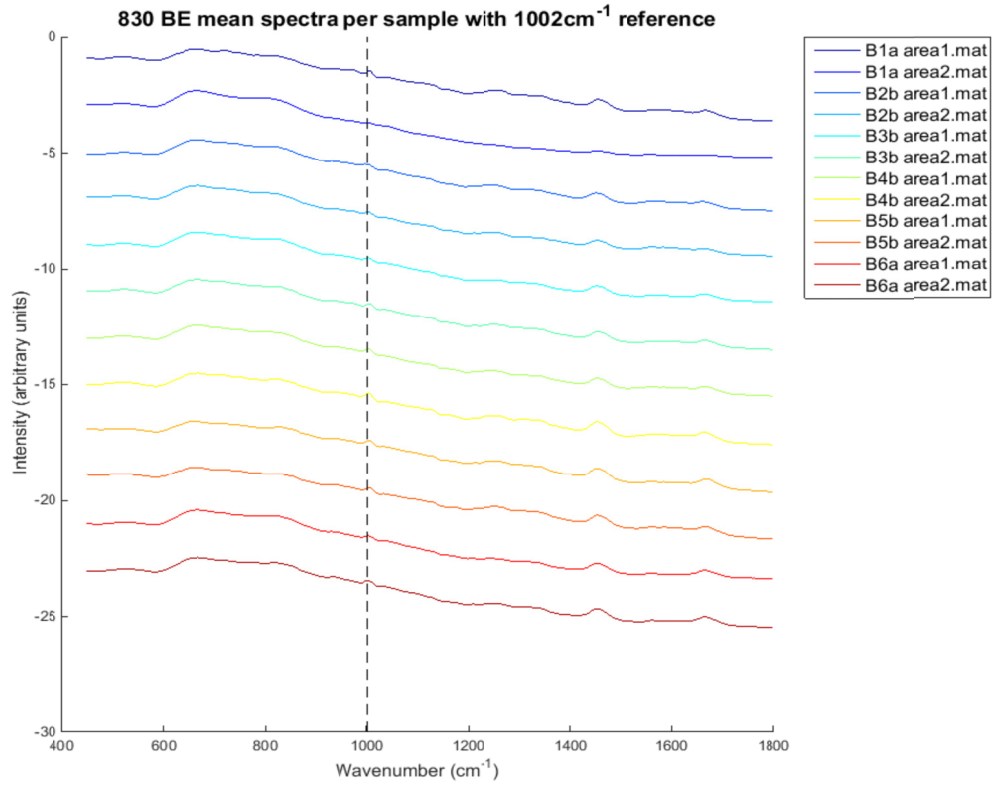
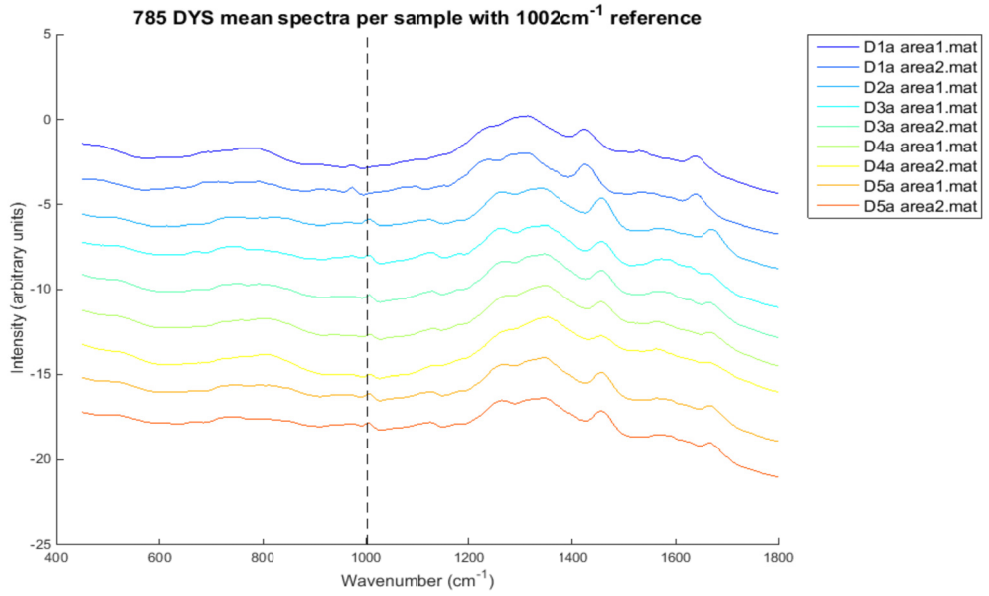


Figure 9-4 Sample mean spectra for Barrett's oesophagus (BE) samples at (A) 785nm and (B) 830nm.

A



B

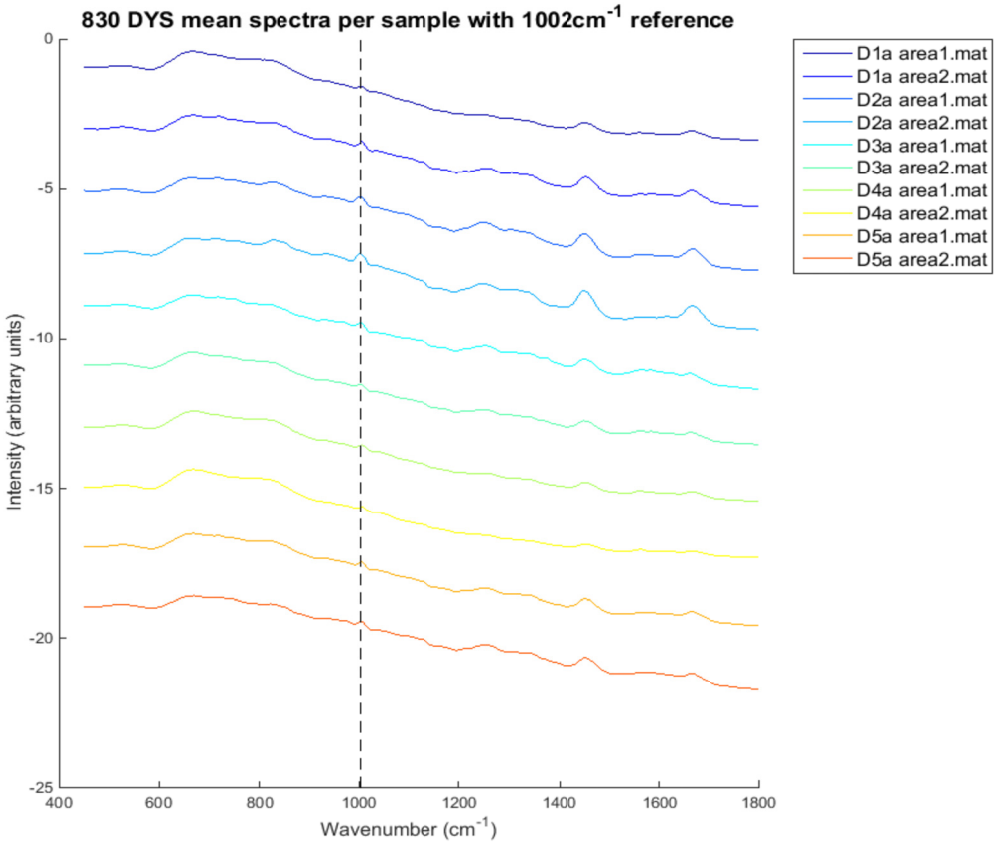
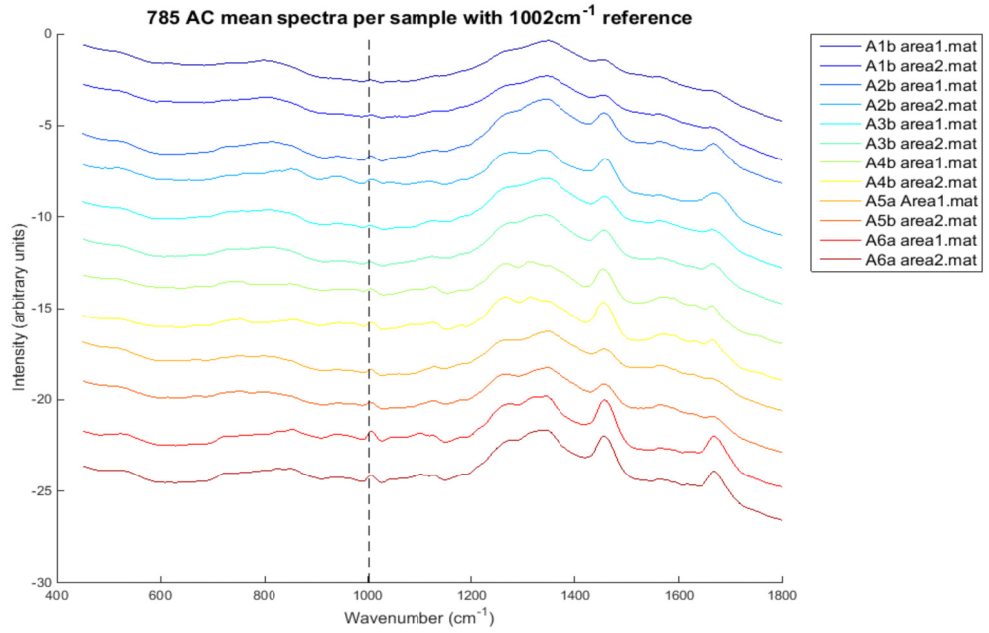


Figure 9-5 Sample mean spectra for dysplasia (DYS) samples at (A) 785nm and (B) 830nm.

A



B

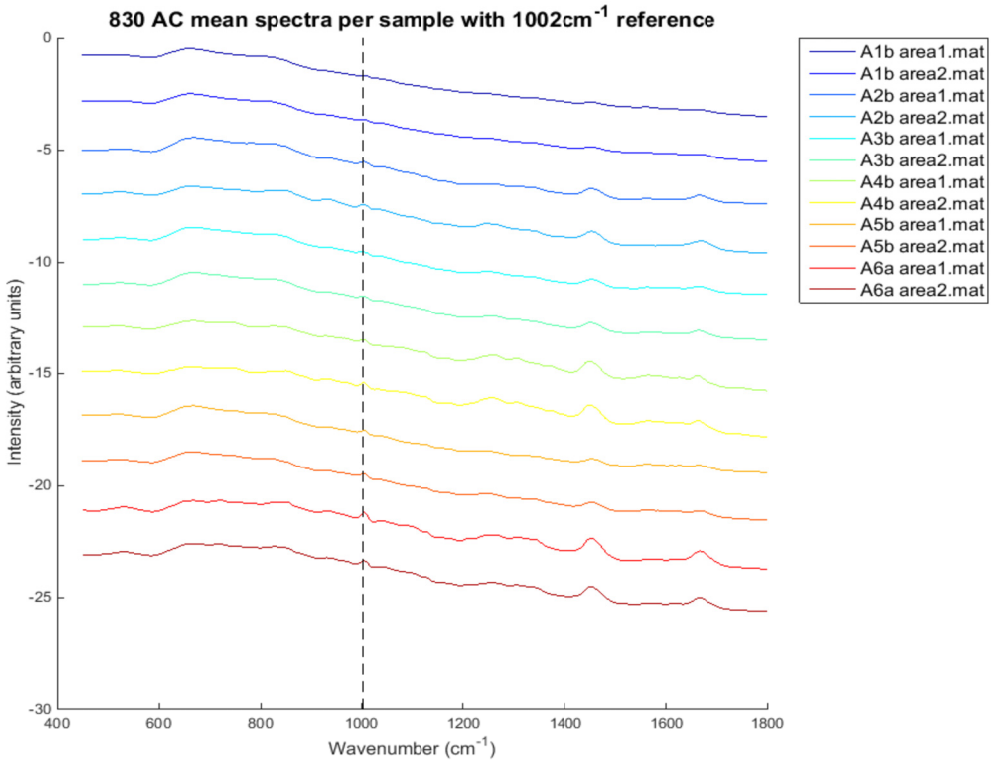


Figure 9-6 Sample mean spectra for adenocarcinoma (AC) samples at (A) 785nm and (B) 830nm.

Looking at the mean spectra it was clear that certain samples showed a peak shift, suggesting problems with the instrument calibration. In the 785nm group, all samples were shifted with a phenylalanine peak around 989cm^{-1}. The Barrett's samples were also shifted, with peak positions between 966-994cm^{-1}. In the dysplasia group, 2 maps from one sample (D1 area 1 and area 2) were significantly shifted at 971-5cm^{-1}, but the others grouped between 1004-1008cm^{-1}. The peak position for every sample in the adenocarcinoma group was 1004cm^{-1}.

At 830nm, one of the normal squamous samples showed phenylalanine peak shift to 988cm^{-1} in all 3 Raman maps (N2 area 1, 2 and 3), with the other normal squamous samples all at 1005cm^{-1}. One Barrett's sample (B2 area 1) was shifted to 997cm^{-1}, with the other peaks all found between 1004-1008cm^{-1}. The dysplasia samples varied between 1000-1008cm^{-1}, and the adenocarcinoma peak positions were from 1000-1005cm^{-1}.

This variability in peak positions had the potential to undermine the basis of spectral classification. Further, the fact that the variation was different across the groups (particularly for the 785nm data), introduced the possibility of artificial differences in the spectra between the groups. Nonetheless, a 'first-look' 3 group classification model was constructed including all the samples, with the results in Table 9-3.

This shows a good classification performance at 785nm, but a poor performance at 830nm. It is quite possible that the sizeable peak shift in many of the samples at 785nm was improving the classification – this would happen if the PCs used this artificial source of variation between the groups to help classify the spectra. At 830nm, the small degree of peak shift in each of the

pathology groups was likely sufficient to impair classification performance, but not systematically different between the groups in a way that would artificially improve separation.

Table 9-3 3 group classification model for all spectra measured at 785nm and 830nm (FTIR results shown for comparison).

3 group classification model comparing 785nm, 830nm and FTIR data				
		NSQ	BE	DYS/AC
785	Sensitivity	97.4	85.5	89.3
	Specificity	94.8	93.1	98.7
830	Sensitivity	77.2	47.8	60.2
	Specificity	95.6	76.1	66.4
IR	Sensitivity	93.6	76.8	84.6
	Specificity	99.9	88.5	85.8

Whatever its effect on performance, the peak shift would be affecting the classification model, and hence ought to be corrected if true biological differences were to form the basis of our classification.

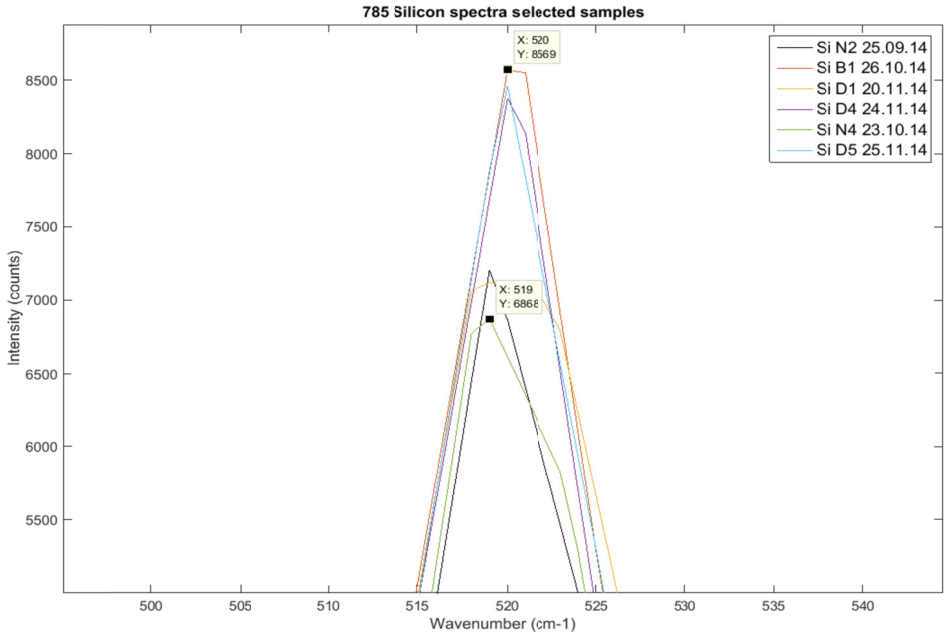
In order to correct for the peak shift a number of approaches were taken. The calibration performed at the time using the silicon peak ought to have corrected for any instrument error. The silicon peak value after the wavenumber correction had been applied was recorded for every calibration: if the instrument

had not been correctly calibrated these values should show the error in the silicon peak. This value could then be applied across all the wavenumbers to correct the peak shift across the entire spectrum.

The silicon peak values recorded prior to measurement of the samples showing peak shift (and certain selected 'unshifted' samples for comparison) are plotted in Figure 9-7.

Figure 9-7 shows that the variation in silicon peak value was at most 2cm^{-1} away from the calibration value of 520.4cm^{-1}. This did not correlate with the magnitude of the variation seen in some of the samples, where the wavenumber shift reached in excess of 30cm^{-1}. This was not an expected result, and we contacted the spectrometer manufacturer for advice. After servicing the instrument they discovered that the calibration process was malfunctioning: when a wavenumber correction was applied to the silicon peak, this correction was then applied by the instrument to the next point measurement. The next point measurement was the test to see if it had been applied correctly (and the calibration value recorded for each map), and this value was therefore correctly calibrated. However this correction was not then applied to the subsequent map as it should have been. Therefore the calibration values we recorded did not reflect the state of the instrument for data acquisition. This was clearly a disappointing finding, and we then had to explore alternative methods to correct the wavenumber shift in our dataset.

A



B

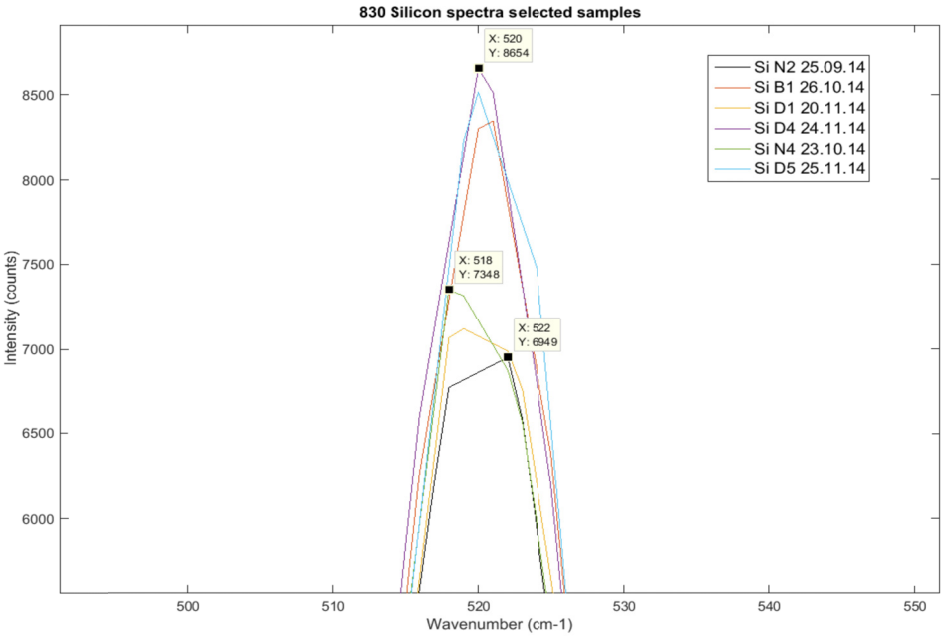


Figure 9-7 Silicon peak values for selected samples showing peak shift and selected 'unshifted' samples for comparison, for (A) 785nm and (B) 830nm. The view shown is a close up of the silicon peak over a small wavenumber range to show the small differences between peak positions.

Using the phenylalanine peak as a reference standard for each individual spectrum, then if its position in that spectrum deviates from 1002cm^{-1} a wavenumber correction can be applied to return the phenylalanine peak to 1002cm^{-1} and correct the entire spectrum (Stone, Kendall, and Barr 2008). We developed an algorithm which calculated the phenylalanine peak position in the mean spectrum for a given sample, then searched for the highest peak value within a +/-5cm^{-1} range of the mean peak position. Each individual spectrum was then shifted to ensure the phenylalanine peak was at 1002cm^{-1}.

The sample means were then re-plotted after peak adjustment, which confirmed that the adjustment had been applied correctly. The peak adjustment for the 785nm Barrett's samples is shown in Figure 9-8. This shows that all of the phenylalanine peaks now align at 1002cm^{-1}, although it does also change the shape of the phenylalanine peak to be a sharper point at a single wavenumber.

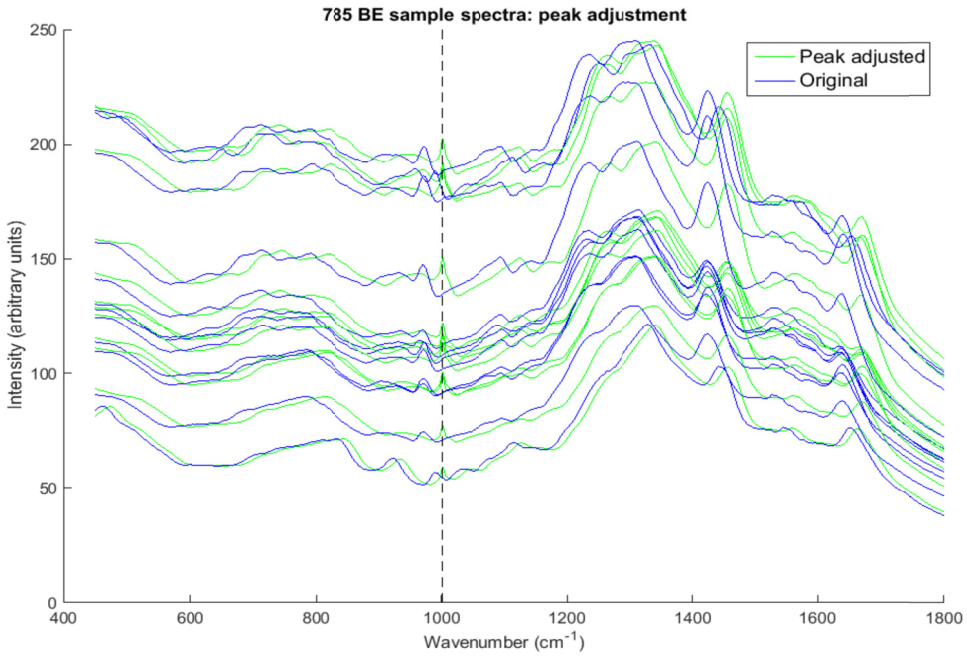
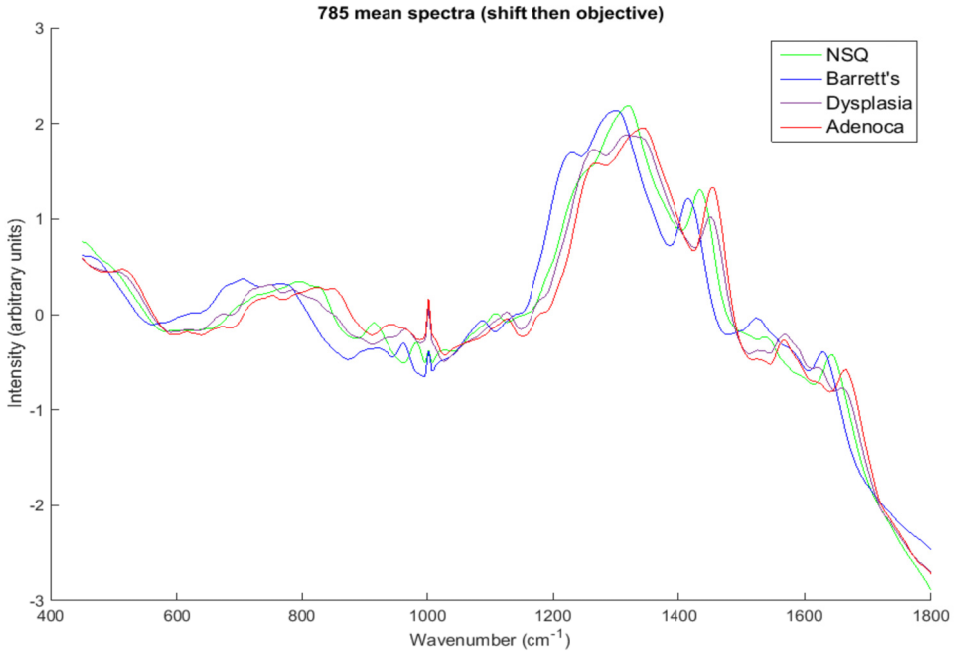


Figure 9-8 Sample mean spectra after peak adjustment for Barrett's samples at 785nm.

A



B

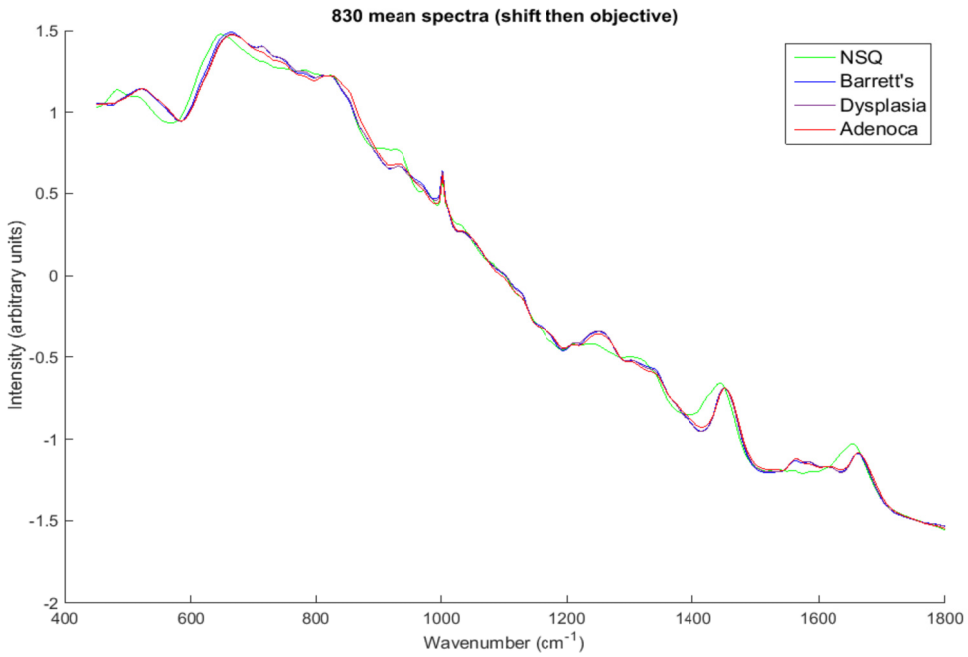


Figure 9-9 Mean spectrum for each pathology group after shift correction, normalisation and objective lens correction, measured at (A) 785nm and (B) 830nm.

After the peak adjustment, it was then necessary to apply the objective lens correction. The spectra after shift and objective lens correction are plotted in Figure 9-9.

This showed that, after peak shift and objective lens correction, although the phenylalanine peaks were all at 1002cm^{-1}, there appeared to be a stretch across the wavenumber range that resulted in many peaks being out of alignment.

We suspected that applying the objective lens correction after the shift correction was causing the stretch effect, and that the objective lens correction algorithm was not appropriate for the peak shifted data. We then looked at the spectra after peak shift only (Figures 9-10A and B). This appeared a much better peak alignment across the entire wavenumber range.

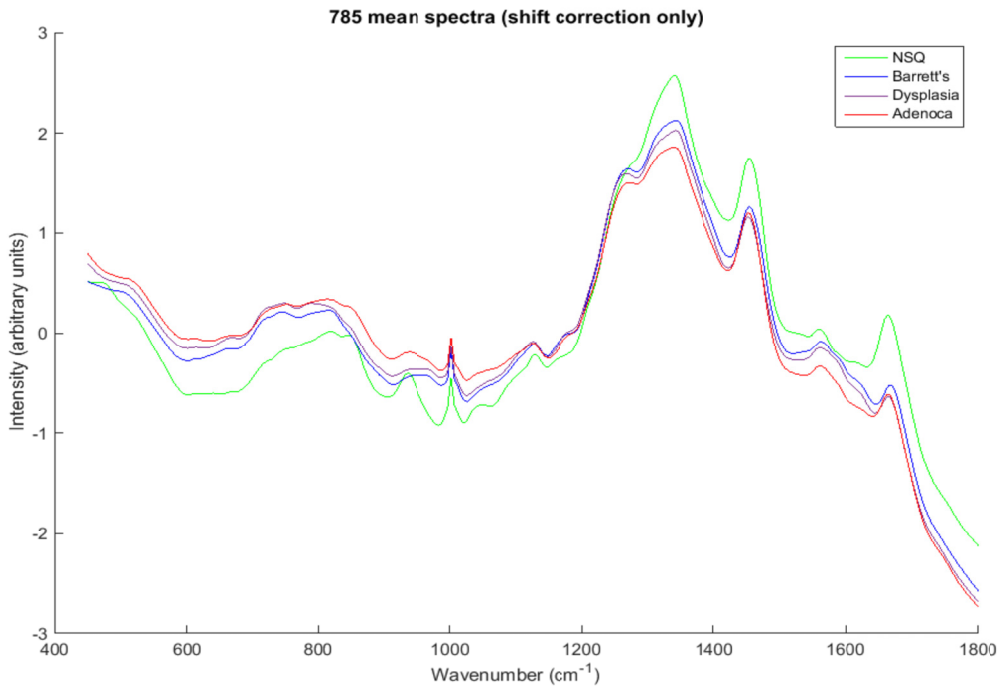


Figure 9-10A Mean spectra after peak shift only, 785nm.

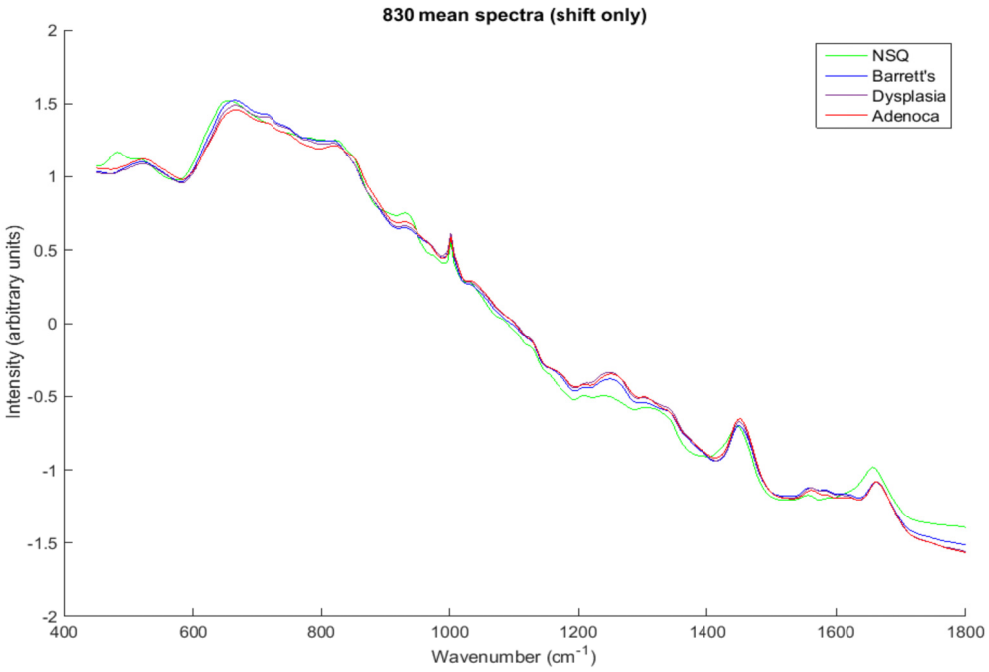


Figure 9-10B Mean spectra after peak shift only, 830nm.

Table 9-4 3 group classification model using shift corrected (but not objective lens corrected) data.

After shift correction but without objective lens correction				
		NSQ	BE	DYS/AC
785	Sensitivity	80.8	72.5	68.0
	Specificity	96.1	79.7	81.5
830	Sensitivity	78.7	60.2	54.8
	Specificity	96.5	69.3	75.9
IR	Sensitivity	93.6	76.8	84.6
	Specificity	99.9	88.5	85.8

The classification performance of the model after shift correction only (with no objective lens correction) is shown in Table 9-4.

This result is better than using the objective lens correction for each pathology group measured at 785nm, and an overall improvement for the 830nm data. Overall however the classification performance is reasonably good for normal squamous tissue, but poor for Barrett's and dysplasia/adenocarcinoma measured at either wavelength.

However, given the difficulties in arriving at this result with the necessary correction algorithms, it is difficult to draw direct comparisons between the 3 modalities as originally intended.

Similarly, spectral peak analysis in this context would not be able to draw any firm conclusions given that all the peaks have been shifted.

Table 9-5 Threshold voting classification using 30% threshold after shift correction (but not objective lens correction). (A) 785nm dataset confusion matrix (B) 785nm dataset classification performance (C) 830nm dataset confusion matrix (D) 830nm classification performance. NSQ normal squamous; BE Barrett's oesophagus; DYS/AC dysplasia/adenocarcinoma

A

	True pathology classification		
Predicted	**NSQ**	**BE**	**DYS/AC**
NSQ	**9**	0	0
BE	0	**7**	2
DYS/AC	2	4	**19**

B

	Tissue pathology		
Performance of 30% voting model	**NSQ**	**BE**	**DYS/AC**
Sensitivity (%)	81.8	63.4	90.5
Specificity (%)	100	93.8	72.7

C

	True pathology classification		
Predicted	**NSQ**	**BE**	**DYS/AC**
NSQ	**9**	0	0
BE	1	**6**	3
DYS/AC	1	5	**19**

D

	Tissue pathology		
Performance of 30% voting model	**NSQ**	**BE**	**DYS/AC**
Sensitivity (%)	81.8	54.5	86.4
Specificity (%)	100	87.9	72.7

To complete the comparison against the FTIR mapping data, these results were then used to construct threshold voting classification of whole samples as explained for FTIR. The results are shown in Table 9-5 above.

These results suggest reasonably good classification of normal squamous and dysplasia/adenocarcinoma groups for both the 785nm and 830nm datasets, although conclusions are limited as discussed above. A summary of these results and comparison against FTIR is shown in Table 9-6 below.

Table 9-6 Comparison of 785nm, 830nm and FTIR mapping for 3 group sample classification model using shift corrected (but not objective lens corrected) data and 30% threshold voting.

Comparison of 785nm, 830nm, and FTIR whole sample classification				
		NSQ	BE	DYS/AC
785	Sensitivity	81.8	63.4	90.5
	Specificity	100	93.8	72.7
830	Sensitivity	81.8	54.5	86.4
	Specificity	100	87.9	72.7
IR	Sensitivity	100	75.0	95.7
	Specificity	100	96.7	86.4

9.8. Discussion

This study aimed to compare the classification performance of Raman mapping at two different excitation wavenumbers: 785nm and 830nm, and enable

comparison with results from mapping the same region using FTIR. The ability of the model to discriminate individual spectra for normal squamous tissue was reasonably good, with sensitivity at 785nm 80.8%, and 78.7% at 830nm. However the classification of Barrett's (sensitivity 72.5% and 60.2% for 785 and 830nm respectively) and particularly dysplasia/adenocarcinoma (sensitivity 68.0% and 54.8% for 785 and 830nm respectively) was poor.

Whole sample classification using a 30% voting threshold improved the ability to detect dysplasia/adenocarcinoma, with 785nm sensitivity 90.5% and 830nm sensitivity 86.4%. At both wavelengths, no 'abnormal' (i.e. Barrett's or dysplasia or adenocarcinoma) sample was classified as normal. However the classification of Barrett's using this system remained poor, with 36.6% of 785nm samples and 45.4% of 830nm samples misclassified as dysplasia /adenocarcinoma.

Our results are consistently better at 785nm than 830nm for all pathology groups. Previous work has shown that 785 and 830nm are the most suitable wavelengths for use in oesophagus (Kendall 2002). The use of a shorter wavelength will give higher tissue signal but a higher autofluorescence background, with a longer wavelength giving a reduced background signal but a weaker tissue signal. Several groups have used Raman in the oesophagus at either 785nm or 830nm, but no comparison of classification has been performed previously. Our results tentatively suggest that 785nm may give slightly better results, however any firm conclusions are severely limited by both the peak shift problem and the strong objective signal seen in this study.

Phenylalanine is a recognised peak that has previously been used as a reference standard (Stone et al. 2008). In the event that the phenylalanine peak

position varied due to genuine differences in tissue biochemistry, then the correction algorithm would have been inappropriate, and a source of error. Whilst the correction algorithm aligned all the phenylalanine peaks successfully, if the underlying calibration problem had caused any stretch as well as shift, this would not have been corrected by the algorithm.

The problems caused by a strong objective lens signal are discussed in Section 8.8 above. The objective lens signal has a high intensity and large peaks for both wavelengths: at 785nm, the objective lens has a large peak at 1100-1500cm^{-1}, with a smaller peak at 700-900cm^{-1}, whereas at 830nm the lower end of the spectrum has a much higher signal intensity, but wavenumbers above 1200cm^{-1} are relatively unaffected. One would expect therefore that the 785nm wavelength may be more suitable for detecting spectral differences in the spectral range up to 1100cm^{-1}, whereas 830nm may perform better for wavenumbers above 1200cm^{-1}. It was not possible to perform this analysis with the peak shift preventing spectral analysis in this study, but this would make an interesting analysis as part of future work.

The objective lens correction algorithm was not appropriate for peak shifted data, and consequently the final analysis of the dataset used peak adjusted data, but without an objective lens correction. The objective lens correction makes adjustments for the varying contribution of the objective lens in samples with a different overall signal intensity, which may be due to tissue fluorescence, or may be affected by other physical properties of the tissue such as thickness. If the correction algorithm is not applied then these physical differences will cause a difference in the amount of signal from the objective lens, and this introduces a further potential source of error.

Many of the same limitations that applied to the FTIR data such as the small sample size, or the difficulties of histology correlation, are true for this dataset, However, the FTIR data demonstrates a level of performance that can be achieved using exactly the same regions. One of the aims of the study was to compare Raman and FTIR for mapping applications, however the problems encountered in the Raman measurements do not provide a fair comparison of optimal measurement parameters.

In comparing between the modalities, it is also possible that the order of measurement may have affected the result. IR is least likely of the modalities to cause photodamage and consequently was measured first, but if there had been tissue damage this could have contributed to the poor performance in the Raman data. If 830nm excitation damaged the tissue, then the 785nm data may have been affected, and this could potentially have impaired the result of its classification model. Given that the performance was better for the 785nm data, however, this seems unlikely.

Proposals for future work as a result of these findings, along with considerations for implementing Raman mapping in clinical practice, are discussed in the summative discussion in Chapter 10.

SECTION D: SUMMATIVE DISCUSSION

CHAPTER 10: SUMMATIVE DISCUSSION AND CONCLUSIONS

10. Summative Discussion and Conclusions

The role of Barrett's oesophagus as the first step in progression towards oesophageal adenocarcinoma has now been recognised for over 50 years. The metaplasia-dysplasia-adenocarcinoma sequence is well characterised, and in this time a wealth of research has informed our understanding of this condition. Developments in endoscopic therapy have provided many tools for advanced endoscopic imaging to aid diagnosis, and revolutionised treatment options, but early diagnosis remains the key to a good outcome. Large-scale surveillance programmes are in place in many countries for patients known to have Barrett's oesophagus: however, such programmes consume a great deal of healthcare resources, and diagnose fewer than 10% of new cases of oesophageal adenocarcinoma.

Improving Barrett's surveillance, along with population screening, are thus major goals of current Barrett's research. The current project used emerging techniques in vibrational spectroscopy to address this problem, with 3 key themes underpinning the various aspects of the project:

i. Development of an automated histopathology tool to reduce the workload of Barrett's surveillance.

ii. Improving detection of dysplasia in Barrett's surveillance.

iii. Development of a non-endoscopic screening tool for Barrett's oesophagus.

Conclusions from this study, progress towards each of these goals, and future work required to take these findings further, are discussed in turn below.

10.1. Development of an automated histopathology tool

This was investigated with both Fourier Transform Infrared (FTIR) and Raman mapping, as described in chapters 4, 8 and 9. The acceptable accuracy of such a tool for translation into clinical practice depends on the exact role required. Our initial aim was to develop a model that could be used for all biopsies taken in the context of a possible diagnosis of Barrett's oesophagus and associated neoplasia. Given the significance of a diagnosis of dysplasia or adenocarcinoma and implications for therapy, it is unlikely that an automated histology result would be considered sufficient diagnostic evidence alone, at least during the introduction of this new technology. Since all diagnoses of dysplasia require dual reporting, the most appropriate position of this technology would be to act as a 'first pathologist' for dysplasia and adenocarcinoma samples. Ideally therefore, accurate classification into 3 groups (normal squamous, Barrett's and dysplasia/adenocarcinoma) is required, to confirm the presence of Barrett's and the need for surveillance, and then identify 'high risk' samples for further review: this was the aim of our 3 group analysis for each modality.

The results with FTIR mapping were promising. Normal squamous samples were identified with 100% sensitivity and specificity by the 3 group model. The sensitivity for dysplasia/adenocarcinoma was 95.6% (with specificity 86.4%), and 75% for the Barrett's group (with specificity 96.7%). The weaker sensitivity of the Barrett's group was partly a result of the threshold voting system chosen to improve sensitivity for dysplasia/adenocarcinoma. In clinical practice this may

be an acceptable trade off, since the samples labelled as ‘dysplastic’ would all have undergone pathology review in our proposed model and should therefore be re-classified correctly. Unnecessary pathology review is a less severe consequence than a missed diagnosis of dysplasia.

A simpler use would be to screen out all the normal samples and review all samples identified as ‘abnormal’ by the model. The 100% sensitivity and specificity for this analysis suggests this can be performed with high accuracy; this role may also be more clinically acceptable as a first step in introducing this new technology.

A different 2 group model could be used to identify ‘high-risk’ samples in patients with known Barrett’s oesophagus: a model well-suited to processing samples taken at surveillance endoscopy. As mentioned earlier, a screening test requires high sensitivity at the expense of low specificity: our 2 group model detected 100% of dysplastic samples with specificity 77.3%. Whilst this false positive rate of 22.7% would mean a relatively high rate of unnecessary pathology review, overall this could dramatically reduce the workload of pathology for Barrett’s surveillance.

If FTIR mapping were to be translated into clinical practice, the major limitations at present are measurement time and cost. Our measurements took 12 minutes per sample but only sampled a tiny fraction of the total area of each tissue section. If used in clinical practice either a random sampling method could be used to cover the whole section, or an automated method for selecting the epithelium would be needed, such as combination with other optical approaches, to ensure an acceptable measurement time. One possible solution has been explored in a variety of cancer models using a combination of

fluorescence spectroscopy with Raman (Kong et al. 2013, 2014). Fluorescence spectroscopy is rapid and highly sensitive but poorly specific: using fluorescence to identify potential areas of disease for subsequent Raman interrogation greatly reduces the time taken.

The current cost of a research FTIR microscope and spectrometer makes this impractical for every pathology unit. However if only a limited number of wavenumbers were required to achieve accurate classification, a modified instrument could be developed which would radically reduce cost and measurement time. Alternatively a more versatile FTIR device which could be used for many other applications may justify a greater cost.

There were a number of limitations to the Raman mapping work investigated in chapters 6 and 7 that contributed to poor classification performance. Chapter 7 aimed to compare Raman wavelengths, and provide a comparison against FTIR mapping, however limited conclusions may be drawn.

For future research applications with biological samples (and if used in clinical practice), one possible remedy for the calibration problems we experienced would be to institute a post measurement check that confirmed the position of phenylalanine peak to within a small margin of error e.g. 998-1006cm^{-1}.

Aside from classification performance, there are several other differences between FTIR and Raman which may affect their use in clinical practice for mapping applications. Chief among these is the time taken for measurement: the FTIR maps took around 12 minutes, whereas the Raman maps took approximately 2 hours. Even if comparable mapping performance were demonstrated, this would likely be a very strong factor in choosing FTIR over Raman if used as a tool for screening pathology samples.

A further factor that may influence choice of modality in clinical practice is the cost – this study was performed using CaF_2 slides which are necessary for Raman and IR in transmission mode, but IR in reflectance mode may be performed more cheaply using e.g. MirrIR slides, which are approximately one tenth the cost of CaF_2 slides.

Although fresh frozen samples were used in this study, techniques for using FTIR on paraffin embedded tissue have been described: this could potentially be a big advantage to reduce the costs and ease tissue processing.

10.2. Improving detection of dysplasia in Barrett's surveillance

Vibrational spectroscopy has the potential to identify dysplastic change at an early stage in the disease process. Neoplastic changes begin with altered gene expression, followed by altered protein content and metabolic effects, then cell atypia or abnormal tissue architecture and then macroscopic abnormalities. It follows that a device capable of detecting the abnormal biochemical changes that occur at a gene/protein/metabolic level could potentially identify neoplastic transformation before cytological or histological changes become apparent. The results in Chapter 5 from FTIR analysis of the 'squamous' cells taken from diseased tissue regions are consistent with the argument that variable phenotypic expression follows genetic and biochemical change.

This is a major strength of the Raman/FTIR approach and could allow earlier detection than conventional cytology/histology for *ex vivo* samples, or other advanced endoscopic systems *in vivo*. However, in the theoretical situation that an ideal diagnostic tool is manufactured and detects changes prior to

histological abnormality, this creates a very challenging situation for testing, since using a gold standard of histology will inevitably be limited.

A combined approach analysing DNA, Raman/FTIR and histology may help to explore this further. This could be performed using multiple contiguous tissue sections from the same samples, with one section stained and examined by a histopathologist, one section measured using Raman/FTIR, and another used for DNA analysis. The DNA analysis could be done using single gland microdissection with whole genome PCR, or investigating specific genomic markers.

Another approach could be to examine the 'outliers', where there is disagreement between Raman/FTIR and histology, and follow up these patients to assess whether the disagreement is due to an 'incorrect' prediction from Raman/FTIR, or whether it is in fact predictive of pathology that is only identifiable at a later stage using conventional histology.

Several strands of this research project explored potential methods to improve the detection of dysplasia over current surveillance techniques of random biopsy and histology. Our attempts to test a novel endoscopic Raman probe are described in Chapter 7. Unfortunately we were not able to carry out the proposed package of experimental work due to the vaporised deposits on the GRIN lenses. The immediate next steps in this project are outlined in chapter 7.

One aim of the comparative Raman mapping study was to inform further probe work with regard to the optimal Raman wavelength. On this evidence 785nm has performed slightly better, but as stated earlier this conclusion is very tentative due to the limitations of this study (discussed in Chapter 9).

Since this study began, significant progress has been made by other groups in developing an endoscopic Raman probe for use in the oesophagus: most notably, a group from Singapore (Bergholt et al. 2014) have published *in vivo* work showing highly promising results. They used a confocal Raman probe (using 785nm) in patients with Barrett's oesophagus, measuring spectra from 373 patients with acquisition times 0.1-0.5s (Bergholt et al. 2014). They report areas under receiver-operator curves (ROC) for dichotomous discrimination as follows: 0.88 for CLO vs. non-dysplastic BO and HGD, 0.84 for non-dysplastic BO vs. CLO and HGD, and 0.90 for HGD vs. non-dysplastic BO and CLO.

A successful endoscopic Raman probe could provide accurate real-time diagnosis at endoscopy. This would allow targeted biopsy to ensure areas of high-grade dysplasia or focal neoplasia were not missed, and facilitate treatment to prevent the progression to oesophageal cancer. It would also enable accurate assessment of the margins of disease *in vivo*, which would ensure completeness of treatment, preventing metastatic spread and minimising the likelihood of radical surgery in the form of oesophagectomy.

Mapping studies also have the potential to improve detection of dysplasia if the techniques could be more sensitive than histology. The study exploring the molecular basis of Raman classification (Chapter 8) aimed to identify key biochemical markers used to discriminate between pathology groups. The results were very limited due to the unexpectedly high signal from the objective lens: this has led to a change in the prototype spectrometer and a lens with a smaller Raman signal will be used. If sensitive biomarkers can be identified, Raman mapping could be used as an adjunct to histology analogous to the role of immunostains.

10.3. Development of a non-endoscopic screening tool for Barrett's oesophagus

Spectral cytopathology (SCP) was investigated as a means of identifying Barrett's oesophagus and associated neoplasia, for potential use in conjunction with a non-endoscopic cell collection device (Chapter 5).

In this study, a reasonably good classification performance was seen for normal squamous samples (sensitivity 79.0%), and dysplasia/adenocarcinoma (83.3%), but the identification of Barrett's samples was very poor (sensitivity 31.3%).

The development of the training dataset was a complex multi-step process with a number of limitations discussed in Chapter 5. However, this result suggests that SCP may be a useful technique for analysing oesophageal cells. There is much potential for optimising the cell preparation process which could potentially greatly enhance the results. Many of these problems related to the small number of glandular cells on the slides, and there were many steps that could be improved, as discussed in Chapter 5.

The results also showed the intriguing possibility that some cells which appear squamous with conventional cytology might show similar biochemical (spectroscopic) changes to those that appear glandular or dysplastic. This would make the technique more robust and less sensitive to difficulties in plating the glandular cells.

Development of an automated cell detection algorithm (the binary mask) was a key component of the study. Further improvement of the binary mask by incorporating a combination of cell size criteria and multiple biochemical peaks could improve this processing step and ensure only cells are detected.

One of the major obstacles to future work in this area is obtaining a reliable gold standard against which to test the training model. The poor sensitivity and specificity in comparable fields (e.g. cervical cytology) suggests that this may limit the usefulness of cytology as a gold standard. Comparing against histology (as in the whole sample test dataset) may be more accurate, although this faces similar limitations

The relatively poor sensitivity and specificity of cytology causes problems for testing, but supports an argument for developing spectral cytopathology since this an area in which diagnostic performance could readily be improved. This technology could potentially be applied in a range of different pathologies and organ systems.

The use of spectral cytopathology in combination with a non-endoscopic collection device could also improve detection of dysplasia: such a device potentially samples the entire length of the oesophagus, and could achieve early detection of dysplasia as described in section 10.2 above.

Whilst this study used samples collected at endoscopy in order to enable histological and endoscopic validation, there may be further challenges using a non-endoscopic cell collection device that collects from the stomach and entire oesophagus. The issue of differentiating pathological glandular cells from stomach cells was not resolved in the present study due to the small numbers of glandular stomach cells. The effectiveness of spectroscopy in discriminating between these groups remains to be seen.

Another major obstacle to clinical implementation is the time needed for sample measurement. In the present study over 20 million spectra were measured using a mapping technique, but only a few thousand were then used in the

training dataset. If a faster automated cell detection process could be used prior to infrared measurement, this would avoid measuring large regions that do not contain cells, and focus solely on collecting useful spectra.

Since the present study began, the first trial data (BEST2) from use of the Cytosponge™ has been published (Ross-Innes et al. 2015). Cases were patients attending for planned endoscopy for Barrett's surveillance, and controls were patients undergoing endoscopy for dyspepsia and/or reflux symptoms. Patients were invited to swallow the Cytosponge™ prior to endoscopy, with endoscopy and biopsy acting as the gold standard for comparison (Barrett's was defined as circumferential Barrett's ≥1cm or maximal length ≥3cm, with intestinal metaplasia on any biopsy). Cytosponge™ samples were processed into paraffin blocks and sections immunostained for TFF3, then scored by a researcher and an expert histopathologist. A total of 1042 patients swallowed the Cytosponge™, with overall sensitivity for detecting Barrett's 79.9%, with this figure increasing to 87.2% in those with ≥3cm of circumferential Barrett's, and a specificity of 92.4%. This is comparable to the sensitivity seen in our study for normal squamous (79.9%) or dysplasia/adenocarcinoma (83.3%). The BEST2 study did not attempt to discriminate dysplasia, but there are plans for future work to incorporate risk stratification using DNA analysis for p53 mutations.

Identification of Barrett's/dysplasia in combination is sufficient if used solely as a screening tool to identify those who require endoscopy. However, reliable detection of dysplasia could potentially replace endoscopic surveillance. The results of our study suggest that this may be achievable with SCP. However, this is a multistep technique which requires further optimisation and validation, which could lead to trials of SCP in combination with a non-endoscopic

collection device such as Cytosponge™. The ability of Cytosponge™ to detect dysplastic Barrett's with p53 risk stratification is not yet known. SCP has the advantages of being automated (whereas immunohistochemistry requires interpretation by a cytopathologist), and, if slide cost were reduced by using transflection mode with MirrIR slides, potentially cheap. However the immunohistochemical and genetic approach may be quicker and more straightforward. Future developments in this field will determine whether either of these techniques is able to fulfil its potential, and revolutionise Barrett's surveillance.

APPENDICES

Appendix I: UICC/AJCC Tumour-Node-Metastasis (TNM) staging system for oesophageal cancer (7th edition).

T Status	
Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	High grade dysplasia
T1	Tumour invades lamina propria, muscularis mucosae or submucosa
T1a	Tumour invades lamina propria or muscularis mucosae
T1b	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour invades adventitia
T4	Tumour invades adjacent strutures
T4a	Resectable tumour invading the pleura, pericardium or diaphragm.
T4b	Unresectable tumour invading other adjacent structures such as the aorta, vertebral body or trachea.

N Status	
Nx	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1-2 regional lymph nodes
N2	Metastasis in 3-6 regional lymph nodes
N3	Metastasis in 7 or more regional lymph nodes

M Status	
Mx	Distant metastases cannot be assessed
M0	No distant metastasis

M1	Distant metastasis

Histological grading	
G1	Well differentiated
G2	Moderately differentiated
G3	Poorly differentiated
G4	Undifferentiated

Adapted from Compton et al. 2012.

Appendix II: Guidelines for surveillance of Barrett's oesophagus

Grade of dysplasia	ACG	ASGE	AGA	BSG	SFED
Non-dysplastic BO	2 OGDs in first year, then every 3 years if no dysplasia	Consider no surveillance. If surveillance chosen, OGD every 3-5 years	OGD every 3-5 years.	BO<3cm (with IM), OGD every 3-5 years. BO≥3cm, OGD every 2-3 years.	BO<3cm, OGD every 5 years. BO3-6cm, OGD every 3 years. BO>6cm, OGD every 2 years.
LGD	Repeat OGD within 6 months; if no HGD, then every 1 year	Repeat OGD within 6 months; if no HGD, then every 1 year	OGD every 6-12 months	Repeat OGD within 3 months; if no HGD, then every 6 months	Repeat OGD. If LGD confirmed, OGD 6 months, 1 year, then yearly.
HGD	Repeat OGD within 3 months, then every 3 months or consider endoscopic therapy.	Consider repeat OGD within 3 months or endoscopic therapy	OGD every 3 months in the absence of endoscopic therapy	Consider endoscopic therapy	Repeat OGD. If HGD confirmed, endoscopic or surgical treatment.

ACG, American College of Gastroenterology; ASGE, American Society for Gastrointestinal Endoscopy; AGA, American Gastroenterological Association; BSG, British Society of Gastroenterology; SFED, French Society of Digestive Endoscopy; BO, Barrett's oesophagus; OGD, oesophago-gastroduodenoscopy; LGD, low-grade dysplasia; HGD high-grade dysplasia. Adapted from de Jonge et al., 2013.

Appendix III: Comparison of advanced endoscopic imaging techniques

Modality type	*Modality*	*Tissue Information*	*Advantages*	*Disadvantages*	*Development stage*
Wide field imaging techniques	High Resolution Endoscopy (HRE)	Topography	Rapid, widely available, simple	Operator dependent, missed lesions	Commercially available
	Chromoendoscopy	Glandular architecture/ vascularity	Low cost, improved lesion detection	Operator dependent, dye application	Commercially available
	Digital chromoendoscopy (e.g. Narrow Band Imaging, NBI)	Glandular architecture/ vascularity	Avoids messy dyes, rapid	Operator dependent	Commercially available
	Autofluorescence Imaging (AFI)	Field change within tissue	Sensitive identification of neoplasia and inflammation	High false positive rate	Commercially available
	Multimodal (HRE/AFI/NBI)	Glandular architecture/ vascularity	Improved detection rate and fewer false positives	Increased cost, operator dependent	Commercially available
	Biomarkers	Molecule-specific	Potentially highly specific at early stage	Suitable biomarkers awaited	In vivo proof of concept
Point measurement techniques	Optical Coherence Tomography	Tissue structure	Defines histology	Insufficient accuracy at present	In vivo feasibility
	Confocal Laser Endomicroscopy	Histology	Defines histology	Operator dependent, slow	Commercially available

	Elastic Scattering Spectroscopy	Cellular structure	Objective, relatively lower cost than other spectroscopy techniques	High signal interference limits accuracy	In vivo feasibility
	Raman Spectroscopy	Biochemical	Objective, defines histology	Weak signal, complex data interpretation	In vivo feasibility

Appendix IV: Consent form for specimen collection

Gloucestershire Hospitals
NHS Foundation Trust
Biophotonics Research
Cranfield Health
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN
Tel 0300 4225470
Fax 0300 4225485
c.kendall@medical-research-centre.com

CONSENT FORM
Version 4 – January 2014

Early Diagnosis of Oesophageal Malignancy

Lead Researchers: Prof H Barr, Dr N Stone, Dr C Kendall

Please initial each box

1. ☐ I confirm that I have read and understand the information sheet, dated January 2014 (version 4), for the above study and have had the opportunity to ask questions.

2. ☐ I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. In the event that I wish to withdraw from this study, data collection will cease immediately and I can request that all data collected is confidentially destroyed.

3. ☐ I understand that Dr N Stone, Dr C Kendall from the Biophotonics Research Group, may access sections of my medical notes, where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.

4. ☐ I agree to take part in this study.

5. ☐ I understand that you may take additional samples during the course of my care to help you develop new diagnostic tests. These samples may be made available for research after all the procedures required for my clinical management have been completed. I give my consent.

…….. Name of Patient	…….. Date	…… Signature
…….. Name of person taking consent (if different from researcher)	…….. Date	…… Signature
…….. Name of Researcher	…….. Date	…… Signature

1 copy for patient, 1 for researcher, 1 for hospital notes

Appendix V: Patient information leaflet

Gloucestershire Hospitals
NHS Foundation Trust
Biophotonics Research
Cranfield Health
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN
Tel 0300 4225470
Fax 0300 4225485
c.kendall@medical-research-centre.com

Early Diagnosis of Oesophageal Malignancy

Lead Researchers: Prof H Barr, Dr N Stone, Dr C Kendall

You are invited to take part in a research study being carried out at the Gloucestershire Royal Hospital. Before you decide whether you would like to participate it is important for you to understand why the research is being done, and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask us if there is anything unclear or if you would like further information. Please take time to decide whether or not you wish to take part. Thank you for reading this information.

What is the purpose of the study?
The staff of Gloucestershire Royal Hospital and scientists at Cranfield University are working together to find new ways of providing advanced clinical support for the local community. An area where we feel we can make major advances is in improving our ability to detect oesophageal disease, pre-cancer and cancer.

Why have I been chosen?
We will be collecting samples from many people under surveillance for Barrett's oesophagus. Therefore this request does not mean that your diagnosis is unsure.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to participate you will be given this information sheet to keep and will be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. Your care, treatment or method of assessment will not be altered in any way by your decision to donate or otherwise.

What will happen to me if I take part?
Your doctor has explained to you that he/she will be removing some oesophagus tissue during the course of your care. Specialist staff within the hospital will examine this tissue and your care guided by what they find. If you are willing to allow us, we can use a small number of additional samples to help us develop new diagnostic tests. No additional procedures will be undertaken as a result of your consent. In addition there is no need for any lifestyle restrictions to participate in this study.

Gloucestershire Hospitals
NHS Foundation Trust
Biophotonics Research
Cranfield Health
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN
Tel 0300 4225470
Fax 0300 4225485
c.kendall@medical-research-centre.com

What is being tested in this study?
We are developing new techniques using light for the early detection of disease. Using advanced rapid tests we hope to be able to easily detect this disease earlier without the need for long uncomfortable procedures. Before these investigations can be carried out on patients, it is necessary to assess its suitability with human tissue in the laboratory. The Biophotonics Research Group based at the Gloucestershire Royal Hospital has over five years experience of working in this field.

What are the possible disadvantages and risks of taking part?
No additional procedures to your routine care will be undertaken as part of this study, therefore there are no associated additional risks.

What are the possible benefits of taking part?
No additional procedures to your routine care will be undertaken as part of this study, therefore there is no immediate clinical benefit to yourself. However, the information collected from this study may help us develop new techniques to improve the treatment of patients with Barrett's oesophagus.

Will my taking part in this study be kept confidential?
All information that is collected about you during the course of the research will be kept strictly confidential. If you consent to taking part in this study you will give permission for Dr N Stone & Dr C Kendall, from the Biophotonics Research Group to have access to your medical notes, where it is relevant to your taking part in research.

What will happen to the results of the research study?
We hope that the results will be published in scientific and medical journals. You will not be identified in any way in such a publication.

Who is funding and organising the research?
The research is funded by the NIHR, The Royal Society & Institute of Physics and Engineering in Medicine. It is being carried out by Cranfield Postgraduate Medical School, a link between the Gloucestershire Hospitals NHS Foundation Trust and Cranfield University.

Who has reviewed the study?
The Gloucestershire Local Research Ethics Committee has reviewed the study.

Consumers for Ethics in Research (CERES) publish a leaflet entitled 'Medical Research and You'. This leaflet gives more information on medical research. A copy can be obtained from CERES, PO Box 1365, London N16 0BW.

Thank you for your help and co-operation. If you have any questions please contact Professor H. Barr, Gloucestershire Royal Hospital. Tel: 0300 4226679 or Dr C. Kendall on 0300 4225470

Appendix VI: Ethical approval for the study

Avon, Gloucestershire and Wiltshire NHS
Health Authority

Gloucestershire LREC
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN
Tel: 01452 395726
Fax: 01452 395720
Email: Hazel.Gage@glouc-r.swest.nhs.uk

Our Ref: 02_112G(28_4)

28 April 2003

Professor Hugh Barr
Consultant Surgeon
Ward 9
Gloucestershire Royal Hospital
Great Western Road
Gloucester GL1 3NN

Dear Professor Barr

Study No 02/112G : Investigation of cancerous and pre-cancerous changes in Barrett's oesophagus using techniques to advance the early diagnosis of malignancy

Thank you for your letter dated 11 April 2003 enclosing the amendments for the above study. I have now reviewed the amendments and will be advising the committee that in my view there is now no objection on ethical grounds to the proposed study. Therefore, I am happy to give you approval on the understanding that you will follow the conditions of the approval set out below. The following documents were reviewed by the committee: -

- LREC application form
- Protocol (Version 2, April 2003)
- Consent Form (Version 1, January 2003)
- Patient Information Sheet (Version 1, January 2003)
- Curriculum Vitae

a) It is the responsibility of the investigator to notify the LREC immediately of any information received by him/her, or of which he/she becomes aware which would cast doubt upon, or alter, any information contained in the original application, a later amendment application or verbal resume submitted to the LREC. The committee should be informed immediately if this information would raise questions about the safety and/or continued conduct of the research.

b) The need to comply with the Data Protection Act 1998.

c) You will need to comply with the Research Governance Framework for Health and Social Care (Department of Health 2001). This includes the requirement to

make available records for the monitoring of the study by the research sponsor. Further information regarding this document can be obtained from Gloucestershire Research & Development Support Unit on 01452 395726.

d) The need to refer proposed amendments to the protocol to the LREC for further review and to obtain LREC approval thereto prior to implementation (except only in cases of emergency where the welfare of the subject is paramount).

e) The requirement to furnish the LREC with details of the progress of the research project periodically (usually annually) and **failure to do this could result in approval to continue with the study being withdrawn.** Please also inform us of the conclusion and outcome of the research project and inform the LREC should the research be discontinued or any subject withdrawn altogether.

f) It is the responsibility of the person conducting any Trial to ensure that all professional staff and management of NHS Trusts involved are notified that it is taking place.

Gloucestershire LREC is fully compliant with the International Conference on Harmonisation/ Good Clinical Practice (ICH GCP) Guidelines for the Conduct of Trials Involving the Participation of Human Subjects.

Please indicate your agreement to comply with the requirements outlined in this letter by signing both copies of this letter and returning one to Hazel Gage. Full approval does not commence until the signed copy is returned.

Yours sincerely

Mike

Dr M J Richards
Chairman, Gloucestershire LREC

cc Dr Sally Pearson, Director of Clinical Strategy, Glos Hospitals NHS Trust
Catherine Kendall, Cranfield Postgraduate Medical School

I agree to comply with the requirements outlined in this letter.

Signed _____ **Date** _____

Gloucestershire Research Ethics Committee
C/o North Bristol NHS Trust
Pembroke Room
Beaufort House
Southmead Hospital
Westbury-on-Trym
Bristol
BS10 5NB

Tel: 0117 323 5211
Fax: 0117 323 2832

14 July 2008

02/112G Investigation of cancerous and pre-cancerous changes in Barrett's oesophagus using techniques to advance the early diagnosis of malignancy

Dear Catherine

Thank you for your letter of 4th July outlining the changes to the study which will involve collaboration with Cancer Research UK to develop the molecular analysis of the samples obtained in the study. I can confirm that Standard Operating Procedures, section 5.3 to 5.40 have been consulted. The committee considers that this alteration is a minor amendment. The amendment does not therefore require ethical review by the committee, and may be implemented immediately provided that it does not affect the research governance approval for the research given by the R & D department for the relevant NHS care organisation.

Documents received

Document	Version Number	Date
Covering letter		4 July 2008
Gloucestershire Committee letter		13 May 2004
Letter from Chief Investigator		6 May 2004
Consent Form	3	July 2008
Patient Information Sheet	3	July 2008

Statement of Compliance

The committee is constituted in accordance with the Governance Arrangements for research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK

Yours sincerely

Anthony Sack
Acting Committee Co-ordinator

anthony.sack@nbt.nhs.uk

The National Research Ethics Service (NRES) represents the NRES Directorate within the National Patient Safety Agency and Research Ethics Committees in England

Appendix VII: Double pathology review of all tissue samples included

(i) Samples included in the FTIR mapping/comparative Raman mapping studies (Chapters 4 and 9).

SAMPLE	PATHOLOGIST 1	PATHOLOGIST 2
N1	NORMAL SQUAMOUS	NORMAL SQUAMOUS
N2	NORMAL SQUAMOUS	NORMAL SQUAMOUS
N3	NORMAL SQUAMOUS	NORMAL SQUAMOUS
N4	NORMAL SQUAMOUS	NORMAL SQUAMOUS
N5	NORMAL SQUAMOUS	NORMAL SQUAMOUS
B1	BARRETT'S	BARRETT'S
B2	BARRETT'S	BARRETT'S
B3	BARRETT'S	BARRETT'S
B4	BARRETT'S	BARRETT'S
B5	BARRETT'S	BARRETT'S
B6	BARRETT'S	BARRETT'S
D1	HGD	LGD
D2	HGD	LGD
D3	HGD	HGD
D4	HGD	LGD

D5	HGD	LGD
A1	ADENOCA.	ADENOCA.
A2	ADENOCA.	ADENOCA.
A3	ADENOCA.	ADENOCA.
A4	ADENOCA.	ADENOCA.
A5	ADENOCA.	ADENOCA.
A6	ADENOCA.	ADENOCA.

(ii) Samples included in the Raman mapping to explore the basis of molecular classification (Chapter 8).

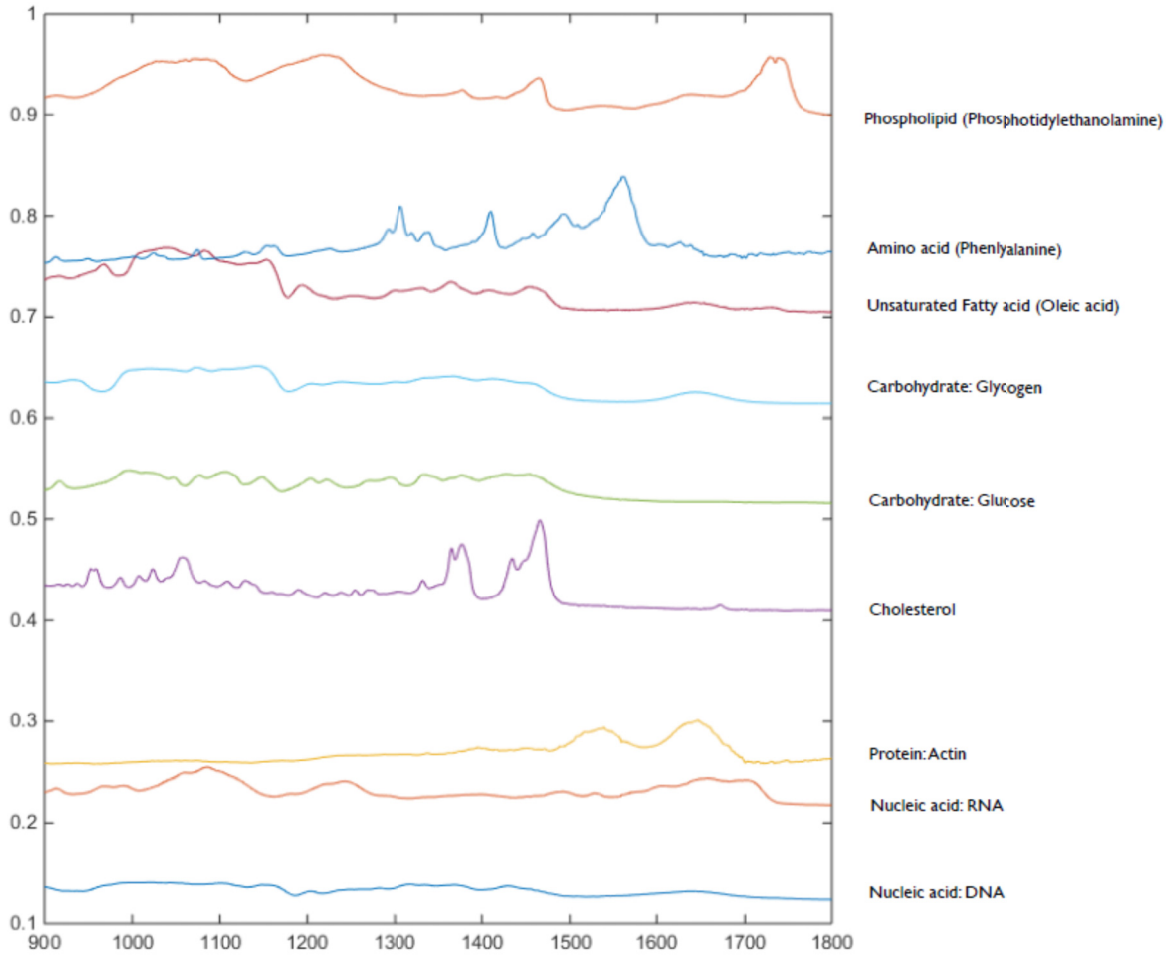
SAMPLE	PATHOLOGIST 1	PATHOLOGIST 2
1	BARRETT'S	BARRETT'S
2	ADENOCA.	MALIGNANT (UNSURE TYPE)
3	BARRETT'S	BARRETT'S
4	BARRETT'S	BARRETT'S
5	BARRETT'S	BARRETT'S
7b	ADENOCA.	ADENOCA.
8a	BARRETT'S	BARRETT'S
9	HGD	INDEFINITE FOR DYSPLASIA

12c	NORMAL SQUAMOUS	NORMAL SQUAMOUS
12d	ADENOCA.	ADENOCA.
14a	ADENOCA.	ADENOCA.
16d	BARRETT'S	BARRETT'S
17	BARRETT'S	BARRETT'S
18	BARRETT'S	BARRETT'S
21	LGD	BARRETT'S
22a1	ADENOCA.	ADENOCA.
23	LGD	BARRETT'S
24X	BARRETT'S	BARRETT'S
24Yb	NORMAL SQUAMOUS	NORMAL SQUAMOUS
25	HGD	LGD
28	NORMAL SQUAMOUS	NORMAL SQUAMOUS
30a	BARRETT'S	BARRETT'S
30b	BARRETT'S	BARRETT'S
32a	HGD	HGD
32b	BARRETT'S	BARRETT'S
33a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
34a	BARRETT'S	BARRETT'S

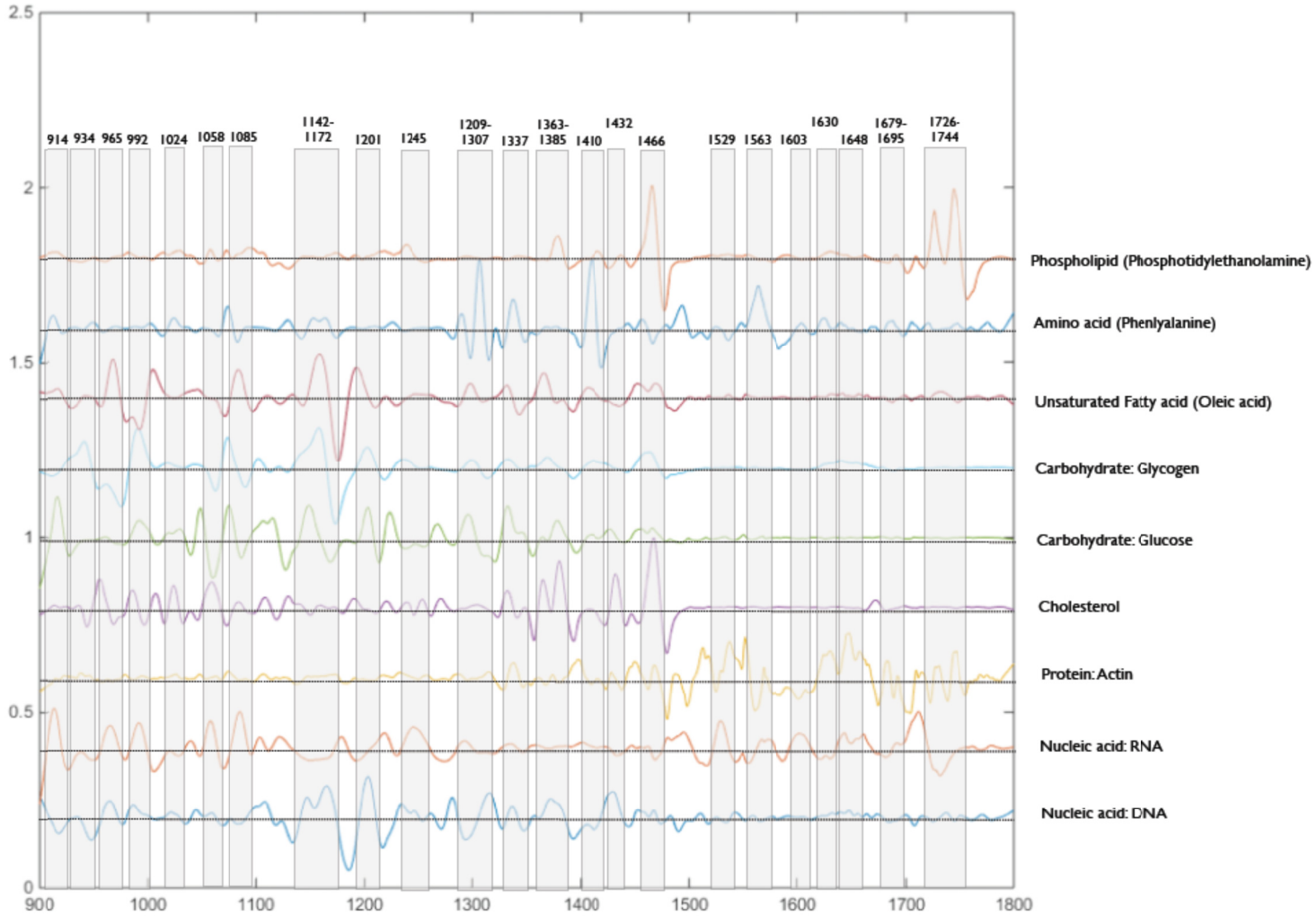
35b	BARRETT'S	BARRETT'S
35c	HGD	HGD
35d	HGD	LGD
36a	ADENOCA.	MALIGNANT (UNSURE TYPE)
37a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
37b	LGD	LGD
38c	NORMAL SQUAMOUS	NORMAL SQUAMOUS
39	HGD	LGD
40	ADENOCA.	ADENOCA.
41b	ADENOCA.	LGD
41c	NORMAL SQUAMOUS	NORMAL SQUAMOUS
43a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
43b	BARRETT'S	BARRETT'S
43c	ADENOCA.	ADENOCA.
45a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
45b	BARRETT'S	BARRETT'S
45c	ADENOCA.	ADENOCA.
46a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
46b	ADENOCA.	LGD

47	LGD	LGD
48a	LGD	LGD
48b	LGD	BARRETT'S
49a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
49b	ADENOCA.	ADENOCA.
50a	NORMAL SQUAMOUS	NORMAL SQUAMOUS
50c	ADENOCA.	ADENOCA.
53b	ADENOCA.	ADENOCA.
54a	ADENOCA.	ADENOCA.
55a	ADENOCA.	MALIGNANT (UNSURE TYPE)
56a	ADENOCA.	ADENOCA.
56c	NORMAL SQUAMOUS	NORMAL SQUAMOUS
57	HGD	HGD
59a	NORMAL SQUAMOUS	NORMAL SQUAMOUS

Appendix VIII: In-house reference spectra



FTIR Biochemical Standards (2nd derivative)



Appendix IX: Experimental plans

1. Evaluation of ‘reproducibility’ of probe measurements
    a. Use polytetrafluoroethylene (PTFE) standard – measure 5 spectra with 3 acquisition times: 0.1s, 1s, 10s.
        i. Measure using different probes (5 ‘new’ probes and compare with ‘old’ prototype probe)
        ii. Measure using different spectrometers (2)
        iii. Measurements by multiple operators (3)
    b. Green glass – repeat as above for different probes and spectrometers
    c. Air background – repeat as above for different probes and spectrometers
2. Repeat for PTFE with 1 acquisition time (1s) for different incident angles (e.g. 15, 30, 45, 60, 75, 90 degrees).
3. Evaluation of probe stability (repeat for each of 5 probes)
    a. Temporal variation – measure PTFE standard with 1s acquisition at 2 minute intervals after switching probe on up to 60 minutes.
    b. Effect of switching laser ‘on’ and ‘off’ after it has warmed up – measure PTFE standard with 1s acquisition after laser switched off for 1 minute x20.
4. Evaluate depth of measurement
    a. Sequential measurements of standard e.g. silicon with increasing layers of PTFE tape (approx. 50 microns thick per layer).
    b. As above but using intra-lipid with varying depth instead of PTFE.
5. Reproducibility in tissue measurements (for 3 covered probes only)

a. Using a tissue sample, measure defined points using a grid – repeat as above for different probes (3)/spectrometers (2)/multiple operators (3).

6. Other variables which may affect probe measurement
    a. Effects on tissue of repeated measurement – measure a defined point using grid, same probe, same spectrometer, same operator, 1s acquisitions x150.
    b. Effect of saline/water/antifoam lavage on fresh tissue measurement – 1s acquisitions x5 with each.
    c. Effect of endoscopic white light/narrow band imaging – x5 1s acquisitions in presence of (i) no ambient light (ii) endoscopic white light (iii) narrow band imaging.
7. Evaluation of diagnostic capability
    a. *Ex vivo* oesophagectomy measurements
        i. Open specimen and wash in distilled water. Secure grid on oesophagectomy specimen, measure spectra at each position (0.1s, 0.5s, 1s acquisition times) then take biopsy from same grid position.
        ii. Unopened oesophagus specimen – passing probe down endoscope with large lumen instrument channel. Measure a spectrum followed immediately by biopsy at each site (can use probe to indent tissue after measurement to identify site for biopsy), repeat 10 sites per specimen.
    b. Endoscopic resection specimens

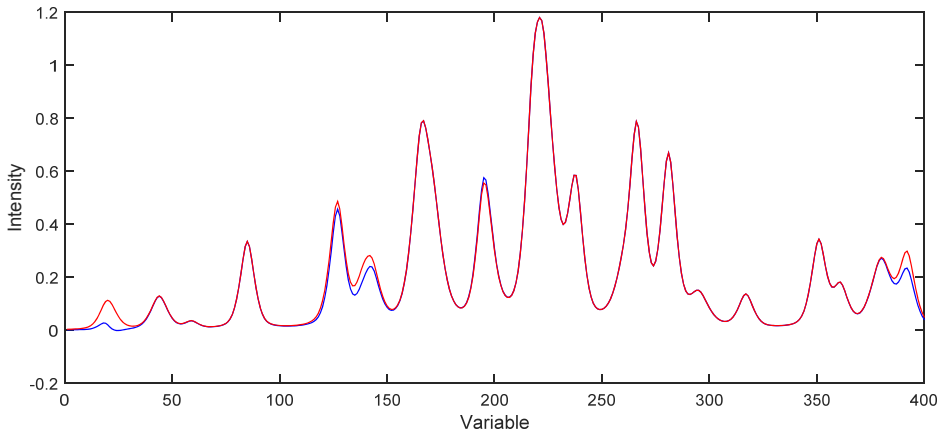
i. Defrost specimen and mount on CaF_2 slide. 0.1s, 0.5s and 1s acquisitions at a number of sites depending on size of specimen.

Appendix X: Objective lens signal simulation

This related work was performed by other members of the Biophotonics team in response to the results measured on the 802 system, presented in Chapter 8. The aim was to use a simulated 2 group dataset and an objective lens with comparable intensity to the measured signal, and model what effect varying the intensity of the objective lens background signal had on the classification ability.

The simulated dataset and objective spectrum is shown in Figure XI-I.

A



B

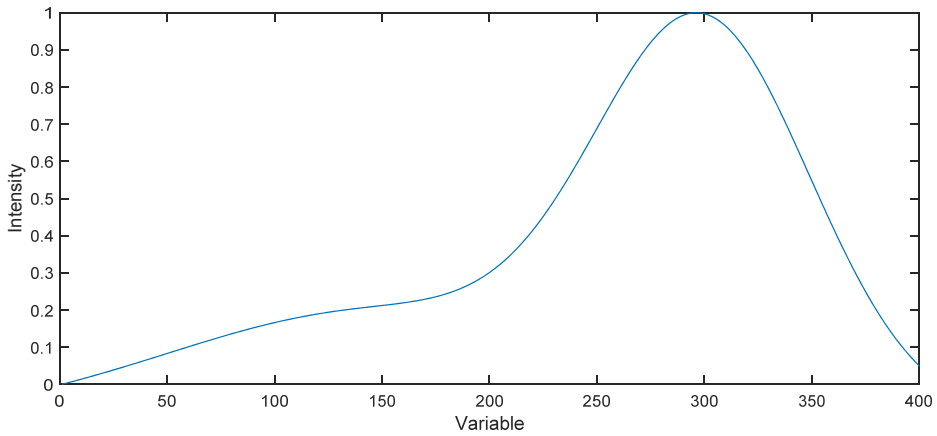


Figure XI-I A Simulated dataset with mean signal from 2 groups. B Objective spectrum

They applied the objective lens signal at varying thresholds to the simulated dataset, then developed a classification model using PCA-fed LDA. The mean spectral signals and associated PCA scatter plots for two values (10% and 70% of the full intensity of the objective lens spectrum) are shown in Figures XI-II A and B.

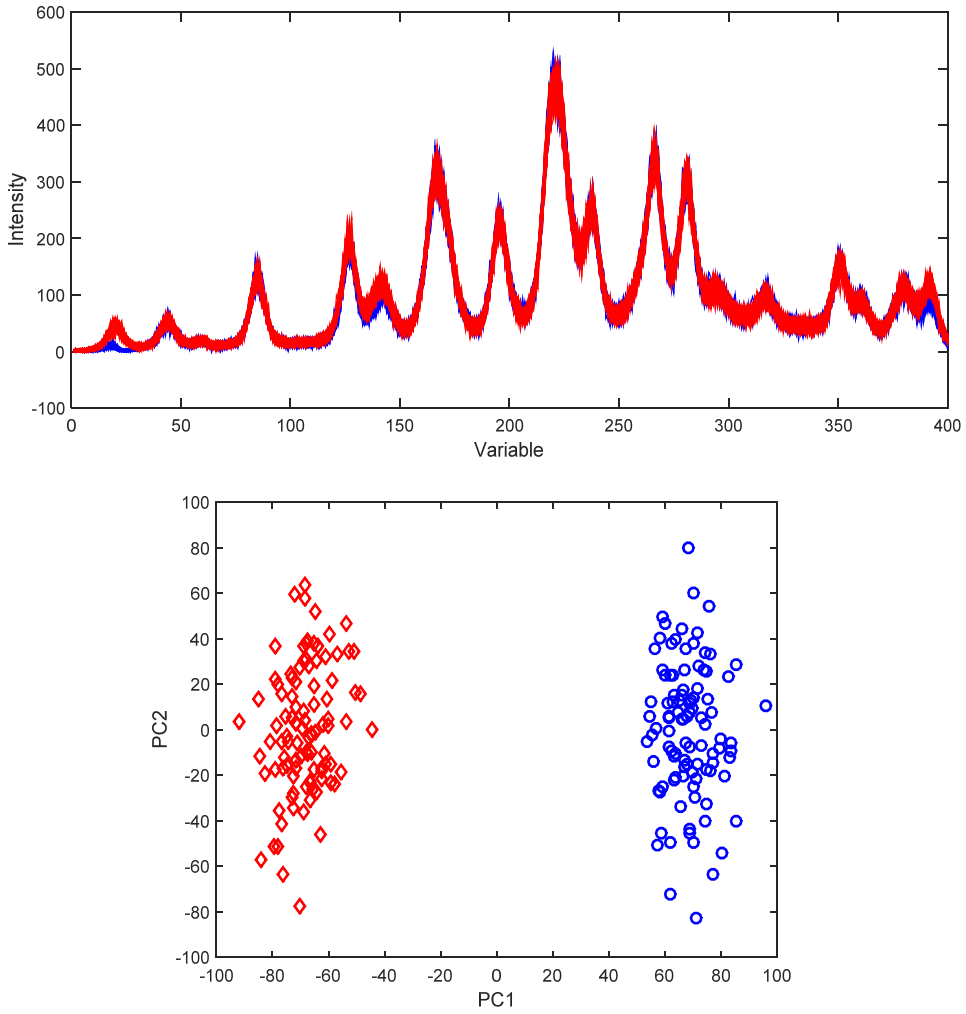


Figure XI-IIA Effect of objective lens signal applied at 10% of full intensity on dataset mean spectra and classification ability on PCA scatter plot.

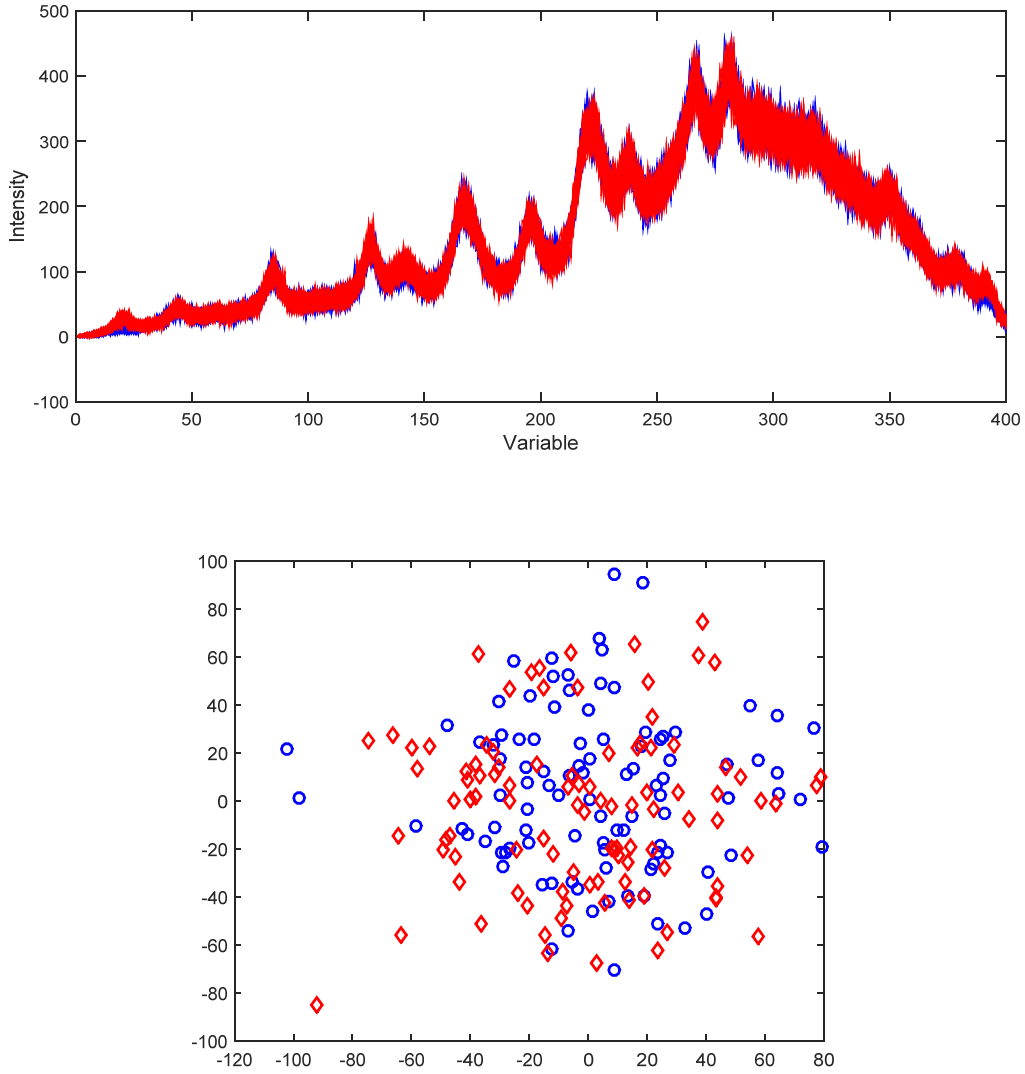


Figure XI-IIB Effect of objective lens signal applied at 70% of full intensity on dataset mean spectra and classification ability on PCA scatter plot.

In Figure XI-IIA, the effect of an objective lens spectrum with 10% of the intensity is seen as a small amount of 'noise' in the signal of the mean spectrum, but PCA can achieve very clear separation of the 2 groups. However at 70% of the intensity (Figure XI-IIB), there is a large increase in the noise in the spectrum, and a complete loss of ability to discriminate on PC score. Figure XI-III shows how the ability of PCA to discriminate varies for objective lens intensity in this simulation.

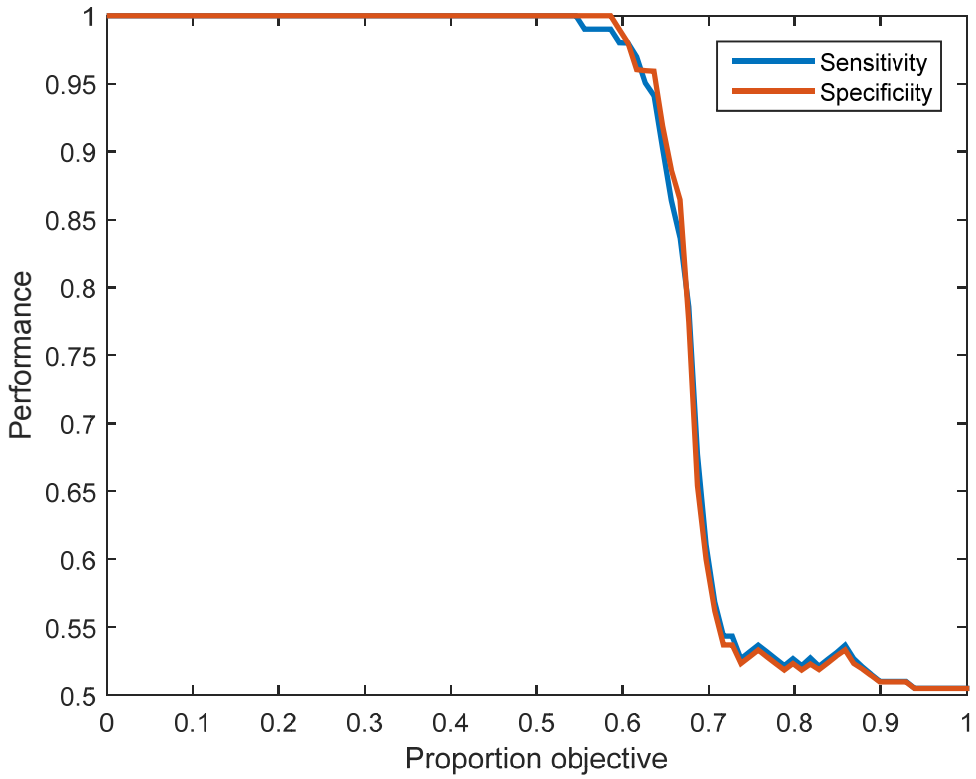


Figure XI-III Effect of objective lens signal intensity ('Proportion objective') on the ability of PCA to discriminate ('Performance') between 2 groups in simulated model.

As the objective intensity increases from 60-75%, the background shot noise overwhelms the differences between the spectra of the two groups. Over multiple iterations, where the classification performance was over 95% accurate with no objective signal present, the discrimination was no better than random (i.e. sensitivity and specificity around 50%) once the objective lens intensity exceeded 75%.

Appendix XI: Prizes, awards, presentations and publications during MD

AWARDS

- Gloucestershire Hospitals NHS Trust Innovation Award 2015: Biophotonics Research Team.
- Stefan & Anna Galeski/Rex & Jean Lawrie Travelling Fellowship, Sri Lanka, March-April 2014.
- University of Exeter Medical School Annual Research Event, 2nd Prize, May 2014.
- BMJ Cancer Team of the Year, Runner up, May 2014.
- British Society of Gastroenterology Poster Prize, June 2014.
- Doyle Travel Award, Heartburn Cancer Awareness and Support, September 2014.
- Harry Morton Anglo-Canadian Travel Fellowship, Royal College of Surgeons Of England/Royal College of Physicians and Surgeons of Canada, October 2014.
- Research Fellowship, Royal College of Surgeons of England, Oct 2013-Oct 2014.

PUBLICATIONS

- Diem M, Miljković M, Bird B, Mazur AI, Schubert JM, Townsend D, Laver N, Almond M, Old O. Cancer screening via infrared spectral cytopathology (SCP): results for the upper respiratory and digestive tracts. Analyst 2015.
- Old O, Moayyedi P, Love S, Roberts C, Hapeshi J, Foy C, Stokes C, Briggs A, Jankowski J, Barr H. 'Barrett's Oesophagus Surveillance

versus endoscopy at need Study (BOSS): protocol and analysis plan for a multicentre randomised controlled trial.' J Med Screen 2015 22(3):158-64

- Old O, Almond M, Barr H. 'Barrett's – how should we manage it?' Frontline Gastroenterol 2015;6:108-116
- D Townsend, M Miljkovic, B Bird, K Lenau, O Old, M Almond, C Kendall, G Lloyd, N Shepherd, H Barr, N Stone, M Diem. Infrared microspectroscopy for cyto-pathological classification of esophageal cells. Analyst 2015 140(7):2215-23
- Old OJ, Hardy T, Kulkarni S, Slim F, Emerson L, Bulbulia R, Whyman M, Poskitt K. 'Incidental non-inguinal hernias in totally extra-peritoneal (TEP) hernia repair. Ann R Coll Surg Engl. 2015 Mar;97(2):120-4
- Almond LM, Old O, Barr H. Strategies for the Prevention of Oesophageal Adenocarcinoma. Int J Surg. 2014;12(9):931-5.
- Weaver JM et al.; the OCCAMS Consortium. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. Nat Genet. 2014;46(8):837-43.
- L. Max Almond, Oliver Old, Nick Stone, Catherine Kendall, Gavin Rhys Lloyd, Joanne Hutchings, John Horsnell, Charlotte Kallaway, Hugh Barr. Real-time disease detection using spectroscopic diagnosis. Biomedical Spectroscopy and imaging, 2014;3(3):197-202
- Rowlands C et al.; on behalf of the Severn and Peninsula Audit and Research Collaborative for Surgeons (SPARCS); the Northwest Research Collaborative. Surgical ward rounds in England: a trainee-led multi-centre study of current practice. Patient Saf Surg. 2014 Feb 28;8(1):11.

- Old, OJ, Fullwood LM, Scott R, Lloyd GR, Almond LM, Shepherd NA, Stone N, Barr H, Kendall C. 'Vibrational Spectroscopy for cancer diagnostics' Anal. Methods, 2014;6:3901-17

Book chapters

- Old OJ, Almond LM, Barr H, Jankowski J. 'Chemoprevention in esophageal carcinoma', in Esophageal cancer Prevention, Diagnosis and Therapy, Springer, 2015.
- Old OJ, Almond LM, Griggs R, Lloyd GR, Horsnell J, Kallaway C, Fullwood L, Das K, Hutchings J, Kendall C, Stone N, Barr H. In vivo optical imaging for early cancer and tumours. OMICS Group e-books, 2014.

PRESENTATIONS

International Presentations

- Old O**.** Early detection of Barrett's dysplasia. St Michael's Hospital, Toronto, October 2014. Invited talk.
- Old O. Vibrational spectroscopy for detection of Barrett's neoplasia. Princess Margaret Cancer Centre, Toronto, October 2014. Invited talk.
- Old OJ, Almond LM, Lloyd GR, Kendall C, Stone N, Barr H. 'Real-time diagnosis of high-grade dysplasia and early oesophageal cancer using Raman spectroscopy'. International Conference of Photodynamic Applications'. Dundee, May 2014.
- Old OJ, Almond LM, Lloyd GR, Kendall C, Stone N, Barr H. 'Real-time diagnosis of high-grade dysplasia and early oesophageal cancer using

Raman spectroscopy'. BioRaman workshop, Warsaw, Poland, May 2014.

- Old OJ, Almond LM, Lloyd GR, Hutchings J, Shepherd N, Barr H, Kendall C, Stone N. Novel fibre-optic Raman probe for real-time diagnosis of early oesophageal cancer. SPEC, Krakow, Poland, August 2014.
- Old O, Almond M, Lloyd G, Townsend D, Lenau K, Diem M, Barr H. Infrared spectral cytopathology as a diagnostic tool in screening for Barrett's esophagus. International Society for Diseases of the Esophagus (ISDE), Vancouver, September 2014. Oral.
- Old O, Stokes C, Woods S, Foy C, Hapeshi J, Barr H; on behalf of the BOSS team. Successful recruitment to Barrett's Oesophagus Surveillance Study (BOSS): you CAN randomize to 'at need' endoscopy. ISDE, Vancouver, September 2014. Oral.
- Old OJ, Almond LM, Lloyd GR, Kendall C, Stone N, Barr H. Advanced endoscopic imaging of Barrett's Esophagus: a feasibility study of Raman Spectroscopy in the presence of endoscopic white light and Narrow Band Imaging. ISDE, Vancouver, September 2014. Oral.

National Presentations

- Raman mapping to classify pathology in Barrett's oesophagus. O Old, M Isabelle, G Lloyd, C Kendall, H Barr, N Stone. CLIRSPEC, Exeter April 2015. Poster.

- Non-endoscopic screening: identifying Barrett's neoplasia with infrared spectroscopy. O Old, LM Almond, G Lloyd, C Kendall, A Shore, H Barr, N Stone. Barrett's Symposium, UCL, April 2015. Poster.
- Dual modality endoscopic therapy in Barrett's associated Oeosphageal Dysplasia in a Tertiary Centre: completing the audit cycle. C J Florance, M Almond, O Old, H Barr. DDF, London, June 2015. Poster.
- Non-endoscopic screening for Barrett's oesophagus: identifying dysplasia with infrared spectroscopy. O Old, LM Almond, G Lloyd, C Kendall, A Shore, H Barr, N Stone. DDF, London, June 2015. Poster.
- Raman mapping to classify pathology in the oesophagus. O Old, M Isabelle, G Lloyd, K Curtis, C Kendall, H Barr, N Stone. DDF, London, June 2015. Poster.
- Selected to present on behalf of RCS Research Fellows at Freemasons, Newport, June 2015.
- Old OJ, Barr H. Barrett's oesophagus: current challenges. Invited speaker, Gastroenterology and Nutrition Symposium, Worcester, September 2014.
- Old OJ, Almond LM, Lloyd GR, Kendall C, Stone N, Barr H. 'Real-time diagnosis of high-grade dysplasia and early oesophageal cancer using Raman spectroscopy'. International Conference of Photodynamic Applications'. Dundee, May 2014. Oral presentation.
- Old O, Stokes C, Woods S, Foy C, Hapeshi J, Barr H; on behalf of the BOSS team. Barrett's Oesophagus Surveillance Study (BOSS): overcoming the challenges of recruitment to a large Randomised Controlled Trial. AUGIS, Brighton, September 2014. Poster.

- Old O, Hardy T, Hewin D, Barr H, Brown J. Post-ERCP Pancreatitis In Secondary Care: Can We Predict Who Would Benefit From A Prophylactic Pancreatic Stent? BSG, Manchester, June 2014.
- Old O, Almond M, Lloyd G, Townsend D, Lenau K, Diem M, Barr H. Barrett's Oesophagus Screening: Infrared Spectroscopy For Cytological Assessment. BSG, Manchester, June 2014. Awarded Poster Prize.
- Old O, Stokes C, Woods S, Foy C, Hapeshi J, Barr H; on behalf of the BOSS team. Barrett's Oesophagus Surveillance Study (BOSS) Update: Successful Recruitment To A Long Follow-up RCT. BSG, Manchester, June 2014.
- Old OJ, Almond LM, Lloyd GR, Diem M, Barr H. 'Infrared spectroscopy for accurate detection of Barrett's oesophagus and dysplasia'. ASGBI, Harrogate, April 2014. Oral presentation.

COURSES

- BEST Academia course: Endoscopic resection and RFA for Barrett's neoplasia. Cambridge, September 2015.
- BioRaman workshop, Warsaw, May 2014
- RCS Research Methods, RCS London, November 2013.
- University of Exeter Professional Skills for Doctors in Specialty Training, Nov 2013-Oct 2014
- Laser Safety Course, Gloucestershire Royal Hospital, December 2013
- Research Grant applications workshop, Taunton, December 2013.

OTHER RESEARCH ACTIVITY

Clinical Trials

- Local Investigator, Gloucestershire Hospitals NHS Foundation Trust, Barrett's Oesophagus Surveillance Study (BOSS): multi-centre randomised controlled trial.
- Local Investigator, Gloucestershire Hospitals NHS Foundation Trust, Aspirin Esomeprazole Chemoprevention Trial (AspECT): multi-centre randomised controlled trial.
- Local Investigator, Gloucestershire Hospitals NHS Foundation Trust, Oesophageal Cancer Clinical and Molecular Strategy (OCCAMS): multi-centre trial.
- Local Investigator, Gloucestershire Hospitals NHS Foundation Trust, Chemoprevention Of Premalignant Intestinal Neoplasia (ChOPIN): multi-centre trial.

Research Supervisor

- I co-designed and supervised a project with a medical student (Avgi Loizidou) from the University of Bristol, who undertook a 4 week special study module in the Biophotonics Research Unit.
- I set up an international collaboration with a group from the University of Toronto. I have co-designed and will co-supervise this project, which will form the basis of a MSc project at the University of Exeter (Emma Upchurch).

BIBLIOGRAPHY

Allison, P. R., and A. S. Johnstone. 1953. “The Oesophagus Lined with Gastric Mucous Membrane.” *Thorax* 8(2):87–101.

Almond, L. M., J. Hodson, and H. Barr. 2014. “Meta-Analysis of Endoscopic Therapy for Low-Grade Dysplasia in Barrett’s Oesophagus.” *The British journal of surgery* 101(10):1187–95.

Almond, L. Max et al. 2011. “Raman Spectroscopy: A Potential Tool for Early Objective Diagnosis of Neoplasia in the Oesophagus.” *Journal of biophotonics* 4(10):685–95.

Almond, L. Max et al. 2012. “Assessment of a Custom-Built Raman Spectroscopic Probe for Diagnosis of Early Oesophageal Neoplasia.” *Journal of biomedical optics* 17(8):081421–1.

Almond, L. Max et al. 2013. “Endoscopic Raman Spectroscopy Enables Objective Diagnosis of Dysplasia in Barrett’s Esophagus.” *Gastrointestinal endoscopy* 4:1–9.

Alvarez Herrero, L. et al. 2010. “Risk of Lymph Node Metastasis Associated with Deeper Invasion by Early Adenocarcinoma of the Esophagus and Cardia: Study Based on Endoscopic Resection Specimens.” *Endoscopy* 42(12):1030–36.

Amano, Yuji et al. 2006. “Which Landmark Results in a More Consistent Diagnosis of Barrett’s Esophagus, the Gastric Folds or the Palisade Vessels?” *Gastrointestinal endoscopy* 64(2):206–11.

Amrania, Hemmel et al. 2012. “Digistain: A Digital Staining Instrument for Histopathology.” *Optics express* 20(7):7290–99.

Anderson, Lesley A. et al. 2007. “Risk Factors for Barrett’s Oesophagus and Oesophageal Adenocarcinoma: Results from the FINBAR Study.” *World journal of gastroenterology : WJG* 13(10):1585–94.

Ashok, Praveen C. et al. 2013. “Multi-Modal Approach Using Raman Spectroscopy and Optical Coherence Tomography for the Discrimination of Colonic Adenocarcinoma from Normal Colon.” *Biomedical optics express* 4(10):2179–86.

Avidan, Benjamin, Amnon Sonnenberg, Thomas G. Schnell, and Stephen J. Sontag. 2002. “Hiatal Hernia and Acid Reflux Frequency Predict Presence and Length of Barrett’s Esophagus.” *Digestive diseases and sciences* 47(2):256–64.

Baker, Matthew J. et al. 2009. “Investigating FTIR Based Histopathology for the Diagnosis of Prostate Cancer.” *Journal of biophotonics* 2(1-2):104–13.

Baker, Matthew J. et al. 2014. “Using Fourier Transform IR Spectroscopy to Analyze Biological Materials.” *Nature protocols* 9(8):1771–91.

Baker, Rebecca et al. 2007. “Depth Profiling of Calcifications in Breast Tissue Using Picosecond Kerr-Gated Raman Spectroscopy.” *Analyst* 132(1):48–53.

Balasubramanian, Gokulakrishnan et al. 2012. “Prevalence and Predictors of Columnar Lined Esophagus in Gastroesophageal Reflux Disease (GERD) Patients Undergoing Upper Endoscopy.” *The American journal of gastroenterology* 107(11):1655–61.

Balu, Mihaela, Gangjun Liu, Zhongping Chen, J. Bruce, and Eric O. Potma. 2010. “Fiber Delivered Probe for Efficient CARS Imaging of Tissues.” *Optics Express* 18(3):2380–88.

Barbour, Andrew P. et al. 2010. “Risk Stratification for Early Esophageal Adenocarcinoma: Analysis of Lymphatic Spread and Prognostic Factors.” *Annals of surgical oncology* 17(9):2494–2502.

Barman, Ishan et al. 2012. “Selective Sampling Using Confocal Raman Spectroscopy Provides Enhanced Specificity for Urinary Bladder Cancer Diagnosis.” *Analytical and bioanalytical chemistry* 404(10):3091–99.

Barrett, N. R. 1950. “Chronic Peptic Ulcer of the Oesophagus and ‘Oesophagitis’.” *The British journal of surgery* 38(150):175–82.

Barrett, N. R. 1957. “The Lower Esophagus Lined by Columnar Epithelium.” *Surgery* 41(6):881–94.

Barth, Andreas. 2007. “Infrared Spectroscopy of Proteins.” *Biochimica et biophysica acta* 1767:1073–1101.

Bassan, Paul et al. 2009. “Resonant Mie Scattering in Infrared Spectroscopy of Biological Materials--Understanding the ‘Dispersion Artefact’.” *The Analyst* 134(8):1586–93.

Bassan, Paul et al. 2010. “Resonant Mie Scattering (RMieS) Correction of Infrared Spectra from Highly Scattering Biological Samples.” *The Analyst* 135(2):268–77.

Bazant-Hegemark, F., K. Edey, G. R. Swingler, M. D. Read, and N. Stone. 2008. “Review: Optical Micrometer Resolution Scanning for Non-Invasive Grading of Precancer in the Human Uterine Cervix.” *Technology in cancer research & treatment* 7(6):483–96.

Beales, Ian, Abigail Hensley, and Yoon Loke. 2013. “Reduced Esophageal Cancer Incidence in Statin Users, Particularly with Cyclo-Oxygenase Inhibition.” *World journal of gastrointestinal pharmacology and therapeutics* 4(3):69–79.

Begley, R. F., AB Harvey, and RL Byer. 1974. “Coherent Anti-Stokes Raman Spectroscopy.” *Applied Physics Letters* 25(7):387.

Bennett, Cathy et al. 2012. “Consensus Statements for Management of Barrett’s Dysplasia and Early-Stage Esophageal Adenocarcinoma, Based on a Delphi Process.” *Gastroenterology* 143(2):336–46.

Bergholt, Mads Sylvest, Wei Zheng, Kan Lin, Khek Yu Ho, Ming Teh, Khay Guan Yeoh, Jimmy Bok Yan So, et al. 2011. “Characterizing Variability in in Vivo Raman Spectra of Different Anatomical Locations in the Upper Gastrointestinal Tract toward Cancer Detection.” *Journal of biomedical optics* 16(3):037003.

Bergholt, Mads Sylvest, Wei Zheng, Kan Lin, Khek Yu Ho, Ming Teh, Khay Guan Yeoh, Jimmy Bok Yan So, et al. 2011. “In Vivo Diagnosis of Gastric Cancer Using Raman Endoscopy and Ant Colony Optimization Techniques.” *International journal of cancer.* 128(11):2673–80.

Bergholt, Mads Sylvest et al. 2013. “Fiber-Optic Raman Spectroscopy Probes Gastric Carcinogenesis in Vivo at Endoscopy.” *Journal of biophotonics* 6(1):49–59.

Bergholt, Mads Sylvest et al. 2014. “Fiberoptic Confocal Raman Spectroscopy for Real-Time in Vivo Diagnosis of Dysplasia in Barrett’s Esophagus.” *Gastroenterology* 146(1):27–32.

Bhargava, Rohit. 2012. “Infrared Spectroscopic Imaging: The next Generation.” *Applied spectroscopy* 66(10):1091–1120.

Bhat, Shivaram et al. 2011. “Risk of Malignant Progression in Barrett’s Esophagus Patients: Results from a Large Population-Based Study.” *Journal of the National Cancer Institute* 103(13):1049–57.

Bhat, Shivaram. 2012. “Barrett’s Oesophagus: Risk of Malignant Progression and Cost Effectiveness of Endoscopic Surveillance.” PhD Thesis, Queen’s University Belfast.

Bird, Benjamin et al. 2008. “Cytology by Infrared Micro-Spectroscopy: Automatic Distinction of Cell Types in Urinary Cytology.” *Vibrational Spectroscopy* 48(1):101–6.

Bird-lieberman, Elizabeth L. et al. 2012. “Molecular Imaging Using Fluorescent Lectins Permits Rapid Endoscopic Identification of Dysplasia in Barrett ’s Esophagus.” *Nature Medicine* 18(2):315-21.

Biscotti, Charles V. et al. 2005. “Assisted Primary Screening Using the Automated ThinPrep Imaging System.” *American Journal of Clinical Pathology* 123(2):281–87.

Bocklitz, T. W., T. Dörfer, R. Heinke, M. Schmitt, and J. Popp. 2015. “Spectrometer Calibration Protocol for Raman Spectra Recorded with Different Excitation Wavelengths.” *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* 149:544–49.

Boerwinkel, D. F. et al. 2014. “Endoscopic TriModal Imaging and Biomarkers for Neoplasia Conjoined: A Feasibility Study in Barrett’s Esophagus.” *Diseases of the esophagus. 27(5):435-43*.

Boerwinkel, D. F. et al. 2014. “Third-Generation Autofluorescence Endoscopy for the Detection of Early Neoplasia in Barrett’s Esophagus: A Pilot Study.” *Diseases of the esophagus*. 27(3):276-84

Borovicka, J. et al. 2009. “Is There an Advantage to Be Gained from Adding Digital Image Cytometry of Brush Cytology to a Standard Biopsy Protocol in Patients with Barrett’s Esophagus?” *Endoscopy* 41(5):409–14.

Botelho, Natalia K. et al. 2010. “Gene Expression Alterations in Formalin-Fixed, Paraffin-Embedded Barrett Esophagus and Esophageal Adenocarcinoma Tissues.” Cancer Biology & Therapy 10(2):172–79.

Boydston-White, Susie et al. 2006. “Cell-Cycle-Dependent Variations in FTIR Micro-Spectra of Single Proliferating HeLa Cells: Principal Component and Artificial Neural Network Analysis.” *Biochimica et biophysica acta* 1758(7):908–14.

Bruun, Susanne W. et al. 2006. “Correcting Attenuated Total Reflection-Fourier Transform Infrared Spectra for Water Vapor and Carbon Dioxide.” *Applied spectroscopy* 60(9):1029–39.

Buskens, Christianne J. et al. 2004. “Prediction of Appropriateness of Local Endoscopic Treatment for High-Grade Dysplasia and Early Adenocarcinoma by EUS and Histopathologic Features.” *Gastrointestinal endoscopy* 60(5):703–10.

Byrne, Hugh J., Ganesh D. Sockalingum, and Nick Stone. 2011. “Raman Microscopy: Complement or Competitor?” in *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach*, edited by David Moss. RSC Analytical Spectroscopy Monographs.

Cameron, Alan J., David J. Ballard, Alan R. Zinsmeister, and J. Aidan Carney. 1990. “Prevalence of Columnar-Lined (Barrett's) Esophagus.” 99:918–22.

Camilleri, Michael et al. 2005. “Prevalence and Socioeconomic Impact of Upper Gastrointestinal Disorders in the United States: Results of the US Upper Gastrointestinal Study.” *Clinical gastroenterology and hepatology* 3(6):543–52.

Camus, Marine et al. 2012. “Helpfulness of the Combination of Acetic Acid and FICE in the Detection of Barrett’s Epithelium and Barrett's Associated Neoplasias.” *World journal of gastroenterology : WJG* 18(16):1921–25.

Cancer Research UK. 2013. “Cancer Statistics.” Retrieved 10 November 2013 (http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/risk/).

Canto, Marcia Irene et al. 2013. “In Vivo Endomicroscopy Improves Detection of Barrett’s Esophagus-Related Neoplasia: A Multicenter International Randomized Controlled Trial (with Video).” *Gastrointestinal endoscopy* 4(2):211–21.

Castanon, A., S. Ferryman, J. Patnick, and P. Sasieni. 2012. “Review of Cytology and Histopathology as Part of the NHS Cervical Screening Programme Audit of Invasive Cervical Cancers.” *Cytopathology* 23(1):13–22.

Chaves, Paula et al. 2007. “Chromosomal Analysis of Barrett’s Cells: Demonstration of Instability and Detection of the Metaplastic Lineage Involved.” *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 20(7):788–96.

Chiriboga, L. et al. 1998. “Infrared Spectroscopy of Human Tissue. II. A Comparative Study of Spectra of Biopsies of Cervical Squamous Epithelium and of Exfoliated Cervical Cells.” *Biospectroscopy* 4(1):55–59.

ClinicalTrials.gov. 2012. “Barrett’s Radiofrequency Intervention for Dysplasia by Endoscopy.” *Clinical Trials.gov*. Retrieved December 19, 2014 (http://clinicaltrials.gov/show/NCT01733719?displayxml=true).

ClinicalTrials.gov. 2013. “A Phase III, Randomized, Study of Aspirin and Esomeprazole Chemoprevention in Barrett’s Metaplasia (AspECT).” Retrieved 15 June 2014 (http://www.clinicaltrials.gov/show/NCT00357682.)

Coad, Rebecca A. et al. 2005. “On the Histogenesis of Barrett’s Oesophagus and Its Associated Squamous Islands: A Three-Dimensional Study of Their Morphological Relationship with Native Oesophageal Gland Ducts.” *The Journal of pathology* 206(4):388–94.

Coco, Dominique P. et al. 2011. “Interobserver Variability in the Diagnosis of Crypt Dysplasia in Barrett Esophagus.” *The American journal of surgical pathology* 35(1):45–54.

Compton, Carolyn C. et al., eds. 2012. *AJCC Cancer Staging Atlas*. New York, NY: Springer New York.

Conrad, Rachel, Camilla Cobb, and Anwar Raza. 2012. “Role of Cytopathology in the Diagnosis and Management of Gastrointestinal Tract Cancers.” *Journal of gastrointestinal oncology* 3(3):285–98.

Cook, M. B., C. P. Wild, and D. Forman. 2005. “A Systematic Review and Meta-Analysis of the Sex Ratio for Barrett’s Esophagus, Erosive Reflux Disease, and Nonerosive Reflux Disease.” *American journal of epidemiology* 162(11):1050–61.

Cooper, Gregory S., Tzyung Doug Kou, and Amitabh Chak. 2009. “Receipt of Previous Diagnoses and Endoscopy and Outcome from Esophageal Adenocarcinoma: A Population-Based Study with Temporal Trends.” *The American journal of gastroenterology* 104(6):1356–62.

Cooper, Gregory S., Zhong Yuan, Amitabh Chak, and Alfred a Rimm. 2002. “Association of Prediagnosis Endoscopy with Stage and Survival in Adenocarcinoma of the Esophagus and Gastric Cardia.” *Cancer* 95(1):32–38.

Corley, Douglas A et al. 2013. “Impact of Endoscopic Surveillance on Mortality from Barrett’s Esophagus-Associated Esophageal Adenocarcinomas.” *Gastroenterology* 145(2):312–19.e1.

Corley, Douglas A. et al. 2007. “Abdominal Obesity and Body Mass Index as Risk Factors for Barrett’s Esophagus.” *Gastroenterology* 133(1):34–41; quiz 311.

Corley, Douglas A., Karla Kerlikowske, Rajiv Verma, and Patricia Buffler. 2003. “Protective Association of aspirin/NSAIDs and Esophageal Cancer: A Systematic Review and Meta-Analysis.” *Gastroenterology* 124(1):47–56.

Corley, Douglas A., Theodore R. Levin, Laurel A. Habel, Noel S. Weiss, and Patricia A. Buffler. 2002. “Surveillance and Survival in Barrett’s Adenocarcinomas: A Population-Based Study.” *Gastroenterology* 122(3):633–40.

Crow, P., A. Molckovsky, et al. 2005. “Assessment of Fiberoptic near-Infrared Raman Spectroscopy for Diagnosis of Bladder and Prostate Cancer.” *Urology* 65(6):1126–30.

Crow, P., B. Barrass, et al. 2005. “The Use of Raman Spectroscopy to Differentiate between Different Prostatic Adenocarcinoma Cell Lines.” *British journal of cancer* 92(12):2166–70.

Curvers, Wouter. L. et al. 2008a. “Endoscopic Tri-Modal Imaging for Detection of Early Neoplasia in Barrett’s Oesophagus: A Multi-Centre Feasibility Study Using High-Resolution Endoscopy, Autofluorescence Imaging and Narrow Band Imaging Incorporated in One Endoscopy System.” *Gut* 57(2):167–72.

Curvers, Wouter et al. 2008b. “Chromoendoscopy and Narrow-Band Imaging Compared with High-Resolution Magnification Endoscopy in Barrett’s Esophagus.” *Gastroenterology* 134(3):670–79.

Curvers, Wouter L., Lorenza Alvarez Herrero, et al. 2010. “Endoscopic Tri-Modal Imaging Is More Effective than Standard Endoscopy in Identifying Early-Stage Neoplasia in Barrett’s Esophagus.” *Gastroenterology* 139(4):1106–14.

Curvers, Wouter L., Fiebo J. ten Kate, et al. 2010. “Low-Grade Dysplasia in Barrett’s Esophagus: Overdiagnosed and Underestimated.” *The American journal of gastroenterology* 105(7):1523–30.

Cuzick, Jack et al. 2009. “Aspirin and Non-Steroidal Anti-Inflammatory Drugs for Cancer Prevention: An International Consensus Statement.” *The lancet oncology* 10(5):501–7.

DaCosta, R. S., H. Andersson, M. Cirocco, N. E. Marcon, and B. C. Wilson. 2005. “Autofluorescence Characterisation of Isolated Whole Crypts and Primary Cultured Human Epithelial Cells from Normal, Hyperplastic, and Adenomatous Colonic Mucosa.” *Journal of clinical pathology* 58(7):766–74.

Das, Kaustuv, Nicholas Stone, Catherine Kendall, Clare Fowler, and J. Christie-Brown. 2006. “Raman Spectroscopy of Parathyroid Tissue Pathology.” *Lasers in medical science* 21(4):192–97.

Dawsey, S. M. et al. 1997. “Studies of Esophageal Balloon Cytology in Linxian, China.” *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 6(2):121–30.

Day, J. C. C. et al. 2009. “A Miniature Confocal Raman Probe for Endoscopic Use.” *Physics in medicine and biology* 54(23):7077–87.

Derenne, Allison, Régis Gasper, and Erik Goormaghtigh. 2011. “The FTIR Spectrum of Prostate Cancer Cells Allows the Classification of Anticancer Drugs according to Their Mode of Action.” *The Analyst* 136(6):1134–41.

Desai, TK, K. Krishnan, N. Samala, and J. Singh. 2012. “The Incidence of Oesophageal Adenocarcinoma in Non-Dysplastic Barrett’s Oesophagus: A Meta-Analysis.” *Gut* 61(7):970–76.

Diem, M., M. Romeo, S. Boydston-White, M. Miljkovic, and C. Matthaus. 2004. “A Decade of Vibrational Micro-Spectroscopy of Human Cells and Tissue (1994-2004).” *The Analyst* 129(10):880–85.

Diem, Max et al. 2015. “Medical Applications of Infrared Spectral Imaging of Individual Cells.” in *Infrared and Raman Spectroscopic Imaging*, edited by R Salzer and HW Siesler.

Diem, Max, Luis Chiriboga, Peter Lasch, and Anthony Pacifico. 2002. “IR Spectra and IR Spectral Maps of Individual Normal and Cancerous Cells.” *Biopolymers* 67(4-5):349–53.

Dingari, Narahara Chari, Gary L. Horowitz, Jeon Woong Kang, Ramachandra R. Dasari, and Ishan Barman. 2012. “Raman Spectroscopy Provides a Powerful Diagnostic Tool for Accurate Determination of Albumin Glycation.” *PLoS ONE* 7(2):e32406.

Dixon, M. F., P. M. Neville, N. P. Mapstone, P. Moayyedi, and A. T. Axon. 2001. “Bile Reflux Gastritis and Barrett’s Oesophagus: Further Evidence of a Role for Duodenogastro-Oesophageal Reflux?” *Gut* 49(3):359–63.

Draga, Ronald O. P. et al. 2010. “In Vivo Bladder Cancer Diagnosis by High-Volume Raman Spectroscopy.” *Analytical chemistry* 82(14):5993–99.

Dulai, Gareth S., Sushovan Guha, Katherine L. Kahn, Jeffrey Gornbein, and Wilfred M. Weinstein. 2002. “Preoperative Prevalence of Barrett’s Esophagus in Esophageal Adenocarcinoma: A Systematic Review.” *Gastroenterology* 122(1):26–33.

Edelstein, Zoe R., Mary P. Bronner, Sheldon N. Rosen, and Thomas L. Vaughan. 2009. “Risk Factors for Barrett’s Esophagus among Patients with

Gastroesophageal Reflux Disease: A Community Clinic-Based Case-Control Study." *The American journal of gastroenterology* 104(4):834–42.

Edelstein, Zoe R., Diana C. Farrow, Mary P. Bronner, Sheldon N. Rosen, and Thomas L. Vaughan. 2007. "Central Adiposity and Risk of Barrett's Esophagus." *Gastroenterology* 133(2):403–11.

Edgren, Gustaf, Hans-Olov Adami, Elisabete Weiderpass Vainio, and Olof Nyrén. 2013. "A Global Assessment of the Oesophageal Adenocarcinoma Epidemic." *Gut* 62(10):1406–14.

Elliott, A., and E. J. Ambrose. 1950. "Structure of Synthetic Polypeptides." *Nature* 165(4206):921–22.

El-Mashtoly, Samir F. et al. 2014. "Automated Identification of Subcellular Organelles by Coherent Anti-Stokes Raman Scattering." *Biophysical journal* 106(9):1910–20.

Eloubeidi, M. A., and D. Provenzale. 2001. "Clinical and Demographic Predictors of Barrett's Esophagus among Patients with Gastroesophageal Reflux Disease: A Multivariable Analysis in Veterans." *Journal of clinical gastroenterology* 33(4):306–9.

El-Serag, Hashem B. et al. 2004. "Proton Pump Inhibitors Are Associated with Reduced Incidence of Dysplasia in Barrett's Esophagus." *The American journal of gastroenterology* 99(10):1877–83.

El-Serag, Hashem B., Peter Kvapil, Joan Hacken-Bitar, and Jennifer R. Kramer. 2005. "Abdominal Obesity and the Risk of Barrett's Esophagus." *The American journal of gastroenterology* 100(10):2151–56.

Endoscopic Classification Review Group. 2005. "Update on the Paris Classification of Superficial Neoplastic Lesions in the Digestive Tract." *Endoscopy* 37(6):570–78.

Estrella, Jeannelyn S. et al. 2011. "Duplicated Muscularis Mucosae Invasion Has Similar Risk of Lymph Node Metastasis and Recurrence-Free Survival as Intramucosal Esophageal Adenocarcinoma." *The American journal of surgical pathology* 35(7):1045–53.

Evans, Conor L. et al. 2005. "Chemical Imaging of Tissue in Vivo with Video-Rate Coherent Anti-Stokes Raman Scattering Microscopy." *Proceedings of the National Academy of Sciences of the United States of America* 102(46):16807–12.

Evans, Conor L., and X. Sunney Xie. 2008. "Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine." *Annual review of analytical chemistry* 1:883–909.

Evans, John A. et al. 2006. "Optical Coherence Tomography to Identify Intramucosal Carcinoma and High-Grade Dysplasia in Barrett's Esophagus." *Clinical gastroenterology and hepatology* 4(1):38–43.

Evans, John A. et al. 2007. “Identifying Intestinal Metaplasia at the Squamocolumnar Junction by Using Optical Coherence Tomography.” *Gastrointestinal endoscopy* 65(1):50–56.

Falk, Gary W. et al. 1997. “Surveillance of Patients With Barrett's Esophagus for Dysplasia and Cancer with Balloon Cytology.” 1787–97.

Falk, Gary W. 2003. “Cytology in Barrett’s Esophagus.” *Gastrointestinal endoscopy clinics of North America* 13(2):335–48.

Fitzgerald, Rebecca C. et al. 2014. “British Society of Gastroenterology Guidelines on the Diagnosis and Management of Barrett’s Oesophagus.” *Gut* 63(1):7–42.

Fléjou, J. F., and M. Svrcek. 2007. “Barrett’s Oesophagus--a Pathologist's View.” *Histopathology* 50(1):3–14.

Fortun, P. J. et al. 2006. “Acetic Acid-Enhanced Magnification Endoscopy in the Diagnosis of Specialized Intestinal Metaplasia, Dysplasia and Early Cancer in Barrett’s Oesophagus.” *Alimentary pharmacology & therapeutics* 23(6):735–42.

Fountoulakis, A. et al. 2004. “Effect of Surveillance of Barrett’s Oesophagus on the Clinical Outcome of Oesophageal Cancer.” *The British journal of surgery* 91(8):997–1003.

Gaddam, Srinivas et al. 2013. “Persistence of Nondysplastic Barrett’s Esophagus Identifies Patients at Lower Risk for Esophageal Adenocarcinoma: Results From a Large Multicenter Cohort.” *Gastroenterology* 145(3):548–53.e1.

Gajjar, Ketan et al. 2012. “Diagnostic Segregation of Human Brain Tumours Using Fourier-Transform Infrared And/or Raman Spectroscopy Coupled with Discriminant Analysis.” *Analytical methods : advancing methods and applications* 5:89–102.

Gajjar, Ketan et al. 2014. “Histology Verification Demonstrates That Biospectroscopy Analysis of Cervical Cytology Identifies Underlying Disease More Accurately than Conventional Screening: Removing the Confounder of Discordance.” *PloS one* 9(1):e82416.

Galmiche, Jean-Paul et al. 2008. “Screening for Esophagitis and Barrett’s Esophagus with Wireless Esophageal Capsule Endoscopy: A Multicenter Prospective Trial in Patients with Reflux Symptoms.” *The American journal of gastroenterology* 103(3):538–45.

Gao, Liang et al. 2012. “Differential Diagnosis of Lung Carcinoma with Coherent Anti-Stokes Raman Scattering Imaging.” *Archives of pathology & laboratory medicine* 136(12):1502–10.

Gao, Yanfeng et al. 2015. “Fourier Transform Infrared Microspectroscopy Monitoring of 5-Fluorouracil-Induced Apoptosis in SW620 Colon Cancer Cells.” *Molecular medicine reports* 11(4):2585–91.

Gatenby, Piers A. C., James R. Ramus, Christine P. J. Caygill, Neil A. Shepherd, and Anthony Watson. 2008. “Relevance of the Detection of Intestinal Metaplasia in Non-Dysplastic Columnar-Lined Oesophagus.” *Scandinavian journal of gastroenterology* 43(5):524–30.

Geisinger, K. R., L. A. Teot, and J. E. Richter. 1992. “A Comparative Cytopathologic and Histologic Study of Atypia, Dysplasia, and Adenocarcinoma in Barrett’s Esophagus.” *Cancer* 69(1):8–16.

Gerson, L. B., R. Edson, P. W. Lavori, and G. Triadafilopoulos. 2001. “Use of a Simple Symptom Questionnaire to Predict Barrett’s Esophagus in Patients with Symptoms of Gastroesophageal Reflux.” *The American journal of gastroenterology* 96(7):2005–12.

Gillen, P., P. Keeling, P. J. Byrne, A. B. West, and T. P. Hennessy. 1988. “Experimental Columnar Metaplasia in the Canine Oesophagus.” *The British journal of surgery* 75(2):113–15.

Goetz, Martin, and Ralf Kiesslich. 2008. “Confocal Endomicroscopy: In Vivo Diagnosis of Neoplastic Lesions of the Gastrointestinal Tract.” *Anticancer research* 28(1B):353–60.

Gora, Michalina J. et al. 2013. “Imaging the Upper Gastrointestinal Tract in Unsedated Patients Using Tethered Capsule Endomicroscopy.” *Gastroenterology* 145(4):723–25.

Grimbergen, M. C. M. et al. 2010. “Signal-to-Noise Contribution of Principal Component Loads in Reconstructed near-Infrared Raman Tissue Spectra.” *Applied spectroscopy* 64(1):8–14.

Grimm, Martin et al. 2010. “MMP-1 Is a (pre-)invasive Factor in Barrett-Associated Esophageal Adenocarcinomas and Is Associated with Positive Lymph Node Status.” Journal of translational medicine 8:99.

Groblewska, Magdalena, Maria Siewko, Barbara Mroczko, and Maciej Szmitkowski. 2012. “The Role of Matrix Metalloproteinases (MMPs) and Their Inhibitors (TIMPs) in the Development of Esophageal Cancer.” Folia Histochemica et Cytobiologica 50(1):12–19.

Gupta, Neil et al. 2012. “Longer Inspection Time Is Associated with Increased Detection of High-Grade Dysplasia and Esophageal Adenocarcinoma in Barrett’s Esophagus.” *Gastrointestinal endoscopy* 76(3):531–38.

Hahn, Hejin P. et al. 2009. “Intestinal Differentiation in Metaplastic, Nongoblet Columnar Epithelium in the Esophagus.” *The American journal of surgical pathology* 33(7):1006–15.

Hammody, Z. et al. 2008. “Distinction of Malignant Melanoma and Epidermis Using IR Micro-Spectroscopy and Statistical Methods.” *The Analyst* 133(3):372–78.

Hardwick, R. H., R. J. Morgan, B. F. Warren, M. Lott, and D. Alderson. 1997. "Brush Cytology in the Diagnosis of Neoplasia in Barrett's Esophagus." *Diseases of the esophagus* 10(4):233–37.

Haris, P. I., and D. Chapman. 1992. "Does Fourier-Transform Infrared Spectroscopy Provide Useful Information on Protein Structures?" *Trends in biochemical sciences* 17(9):328–33.

Harris, Andrew T. et al. 2010. "Raman Spectroscopy in Head and Neck Cancer." *Head & neck oncology* 2:26.

Harrison, Rebecca et al. 2007. "Detection of Intestinal Metaplasia in Barrett's Esophagus: An Observational Comparator Study Suggests the Need for a Minimum of Eight Biopsies." *The American journal of gastroenterology* 102(6):1154–61.

Harvard University Department of Chemistry and Chemical Biology. 2015. "Coherent Raman Imaging." *Harvard University Department of Chemistry and Chemical Biology*. Retrieved 15 March 2015 (http://bernstein.harvard.edu/research/cars-why.htm).

Herszenyi, Laszlo et al. 2007. "Alteration of Glutathione S-Transferase and Matrix Metalloproteinase-9 Expressions Are Early Events in Esophageal Carcinogenesis." World Journal of Gastroenterology 13(5):676–82.

Huang, Zhiwei et al. 2003. "Near-Infrared Raman Spectroscopy for Optical Diagnosis of Lung Cancer." *International journal of cancer. Journal international du cancer* 107(6):1047–52.

Hughes, Jonathan H., D. Ph, and Michael B. Cohen. 1998. "Is the Cytologic Diagnosis of Esophageal Glandular Dysplasia Feasible ?" *Diagnostic cytopathology* 18(4):312–16.

Hutchings, Joanne et al. 2009. "The Potential for Histological Screening Using a Combination of Rapid Raman Mapping and Principal Component Analysis." *Journal of biophotonics* 2(1-2):91–103.

Hutchings, Joanne, Catherine Kendall, Neil Shepherd, Hugh Barr, and Nicholas Stone. 2010. "Evaluation of Linear Discriminant Analysis for Automated Raman Histological Mapping of Esophageal High-Grade Dysplasia." *Journal of biomedical optics* 15(6):066015.

Hvid-Jensen, F., and L. Pedersen. 2011. "Incidence of Adenocarcinoma among Patients with Barrett's Esophagus." *New England Journal of Medicine* 365(15):1375–83.

Isabelle, M., N. Stone, and H. Barr. 2008. "Lymph Node Pathology Using Optical Spectroscopy in Cancer Diagnostics." *Spectroscopy* 22(2008):97–104.

Ito, Hiroaki et al. 2013. "Use of Surface-Enhanced Raman Scattering for Detection of Cancer-Related Serum-Constituents in Gastrointestinal Cancer

Patients." *Nanomedicine : nanotechnology, biology, and medicine* 10(3):599-608.

Jankowski, J. a et al. 1999. "Molecular Evolution of the Metaplasia-Dysplasia-Adenocarcinoma Sequence in the Esophagus." *The American journal of pathology* 154(4):965–73.

Jankowski, Janusz A., Dawn Provenzale, and Paul Moayyedi. 2002. "Esophageal Adenocarcinoma Arising from Barrett's Metaplasia Has Regional Variations in the West." *Gastroenterology* 122(2):588–90.

Jankowski, Janusz, Hugh Barr, Ken Wang, and Brendan Delaney. 2010. "Diagnosis and Management of Barrett 's Oesophagus." *British Medical Journal* 341:c4551.

Ji, Minbiao et al. 2013. "Rapid, Label-Free Detection of Brain Tumors with Stimulated Raman Scattering Microscopy." *Science translational medicine* 5(201):201ra119.

Jimenez-Hernandez, M. et al. 2013. "Exploring the Spectroscopic Differences of Caki-2 Cells Progressing through the Cell Cycle While Proliferating in Vitro." *The Analyst* 138(14):3957–66.

Jobe, Blair A. et al. 2006. "Office-Based Unsedated Small-Caliber Endoscopy Is Equivalent to Conventional Sedated Endoscopy in Screening and Surveillance for Barrett's Esophagus: A Randomized and Blinded Comparison." *The American journal of gastroenterology* 101(12):2693–2703.

De Jonge, Pieter J. F. et al. 2010. "Risk of Malignant Progression in Patients with Barrett's Oesophagus: A Dutch Nationwide Cohort Study." *Gut* 59(8):1030–36.

De Jonge, Pieter Jan F., Mark van Blankenstein, William M. Grady, and Ernst J. Kuipers. 2013. "Barrett's Oesophagus: Epidemiology, Cancer Risk and Implications for Management." *Gut* 2013.

Juhasz, Arpad et al. 2011. "Prevalence of Barrett Esophagus in First-Degree Relatives of Patients with Esophageal Adenocarcinoma." *Journal of clinical gastroenterology* 45(10):867–71.

Kadri, Sudarshan R. et al. 2010. "Acceptability and Accuracy of a Non-Endoscopic Screening Test for Barrett's Oesophagus in Primary Care: Cohort Study." *BMJ (Clinical research ed.)* 341:c4372.

Kahrilas, Peter J. 2011. "The Problems with Surveillance of Barrett's Esophagus." *The New England journal of medicine* 365(15):1437–38.

Kara, M. A et al. 2007. "Characterization of Tissue Autofluorescence in Barrett's Esophagus by Confocal Fluorescence Microscopy." *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / I.S.D.E* 20(2):141–50.

Kara, Mohammed A, Femke P. Peters, Paul Fockens, Fiebo J. W. ten Kate, and Jacques J. G. H. M. Bergman. 2006. “Endoscopic Video-Autofluorescence Imaging Followed by Narrow Band Imaging for Detecting Early Neoplasia in Barrett’s Esophagus.” *Gastrointestinal endoscopy* 64(2):176–85.

Kara, Mohammed A. et al. 2005. “Endoscopic Video Autofluorescence Imaging May Improve the Detection of Early Neoplasia in Patients with Barrett’s Esophagus.” *Gastrointestinal Endoscopy* 61(6):679–85.

Kara, Mohammed, Ralph S. DaCosta, Brian C. Wilson, Norman E. Marcon, and Jacques Bergman. 2004. “Autofluorescence-Based Detection of Early Neoplasia in Patients with Barrett’s Esophagus.” *Digestive diseases (Basel, Switzerland)* 22(2):134–41.

Kast, R. E. et al. 2007. “Raman Spectroscopy Can Differentiate Malignant Tumors from Normal Breast Tissue and Detect Early Neoplastic Changes in a Mouse Model.” *Biopolymers* 89(3):235–41.

Kastelein, Florine et al. 2013. “Aberrant p53 Protein Expression Is Associated with an Increased Risk of Neoplastic Progression in Patients with Barrett’s Oesophagus.” *Gut* 62(12):1676–83.

Kaye, Philip V et al. 2009. “Barrett’s Dysplasia and the Vienna Classification: Reproducibility, Prediction of Progression and Impact of Consensus Reporting and p53 Immunohistochemistry.” *Histopathology* 54(6):699–712.

Kaye, Philip V et al. 2010. “Novel Staining Pattern of p53 in Barrett’s Dysplasia--the Absent Pattern.” *Histopathology* 57(6):933–35.

Kazarian, Sergei G., and K. L. Andrew Chan. 2013. “ATR-FTIR Spectroscopic Imaging: Recent Advances and Applications to Biological Systems.” *The Analyst* 138(7):1940–51.

Keld, R. et al. 2010. “The ERK MAP Kinase-PEA3/ETV4-MMP-1 Axis Is Operative in Oesophageal Adenocarcinoma.” Mol Cancer 9:313.

Kendall, Catherine. 2002. “A Study of Raman Spectroscopy for the Early Detection and Classification of Malignancy in Oesophageal Tissue.” Cranfield University.

Kendall, Catherine et al. 2003. “Raman Spectroscopy, a Potential Tool for the Objective Identification and Classification of Neoplasia in Barrett’s Oesophagus.” *The Journal of pathology* 200(5):602–9.

Kendall, Catherine et al. 2009. “Vibrational Spectroscopy: A Clinical Tool for Cancer Diagnostics.” *The Analyst* 134(6):1029–45.

Kerkhof, M. et al. 2007. “Grading of Dysplasia in Barrett’s Oesophagus: Substantial Interobserver Variation between General and Gastrointestinal Pathologists.” *Histopathology* 50(7):920–27.

Khan, S. et al. 1998. “Diagnostic Value of p53 Immunohistochemistry in Barrett’s Esophagus: An Endoscopic Study.” *Pathology* 30(2):136–40.

Kitchener, Henry C. et al. 2011. “Automation-Assisted versus Manual Reading of Cervical Cytology (MAVARIC): A Randomised Controlled Trial.” *The Lancet. Oncology* 12(1):56–64.

Kohler, Achim et al. 2008. “Estimating and Correcting Mie Scattering in Synchrotron-Based Microscopic Fourier Transform Infrared Spectra by Extended Multiplicative Signal Correction.” *Applied spectroscopy* 62(3):259–66.

Konda, Vani J. A. et al. 2008. “Is the Risk of Concomitant Invasive Esophageal Cancer in High-Grade Dysplasia in Barrett’s Esophagus Overestimated?” *Clinical gastroenterology and hepatology* 6(2):159–64.

Kong, Kenny et al. 2013. “Diagnosis of Tumors during Tissue-Conserving Surgery with Integrated Autofluorescence and Raman Scattering Microscopy.” *Proceedings of the National Academy of Sciences of the United States of America* 110(38):15189–94.

Kong, Kenny et al. 2014. “Towards Intra-Operative Diagnosis of Tumours during Breast Conserving Surgery by Selective-Sampling Raman Micro-Spectroscopy.” *Physics in medicine and biology* 59(20):6141–52.

Kong, Kenny, Catherine Kendall, Nicholas Stone, and Ioan Notingher. 2015. “Raman Spectroscopy for Medical Diagnostics - From in-Vitro Biofluid Assays to in-Vivo Cancer Detection.” *Advanced Drug Delivery Reviews*.89:121-34

Krafft, Christoph, Anuradha a Ramoji, et al. 2009. “A Comparative Raman and CARS Imaging Study of Colon Tissue.” *Journal of biophotonics* 2(5):303–12.

Krafft, Christoph. 2009. “Vibrational Spectroscopic Imaging of Soft Tissues.” Pp. 113–47 in *Infrared and Raman Spetroscopic Imaging*, edited by Reiner Salzer and Heinz Siesler. Weinheim: Wiley-VCH.

Krafft, Christoph, Benjamin Dietzek, and Jürgen Popp. 2009. “Raman and CARS Microspectroscopy of Cells and Tissues.” *The Analyst* 134(6):1046–57.

Krafft, Christoph, and Valter Sergo. 2006. “Biomedical Applications of Raman and Infrared Spectroscopy to Diagnose Tissues.” *Spectroscopy* 20:195–218.

Kubo, Ai et al. 2009. “Cigarette Smoking and the Risk of Barrett’s Esophagus.” *Cancer causes & control : CCC* 20(3):303–11.

Labenz, J., and P. Malfertheiner. 1997. “Helicobacter Pylori in Gastro-Oesophageal Reflux Disease: Causal Agent, Independent or Protective Factor?” *Gut* 41(3):277–80.

Lagarde, S. M., F. J. W. ten Kate, D. J. Richel, G. J. A. Offerhaus, and J. J. B. van Lanschot. 2007. “Molecular Prognostic Factors in Adenocarcinoma of the Esophagus and Gastroesophageal Junction.” Annals of surgical oncology 14(2):977–91.

Lagergren, J., R. Bergström, A. Lindgren, and O. Nyrén. 1999. “Symptomatic Gastroesophageal Reflux as a Risk Factor for Esophageal Adenocarcinoma.” *The New England journal of medicine* 340(11):825–31.

Lao-Sirieix, P. et al. 2009. “Non-Endoscopic Screening Biomarkers for Barrett’s Oesophagus: From Microarray Analysis to the Clinic.” *Gut* 58(11):1451–59.

Lasch, Peter, Wolfgang Haensch, Dieter Naumann, and Max Diem. 2004. “Imaging of Colorectal Adenocarcinoma Using FT-IR Microspectroscopy and Cluster Analysis.” *Biochimica et biophysica acta* 1688(2):176–86.

Lavery, Danielle L. et al. 2014. “The Stem Cell Organisation, and the Proliferative and Gene Expression Profile of Barrett’s Epithelium, Replicates Pyloric-Type Gastric Glands.” *Gut* 63(12):1854–63.

Leggett, Cadman L., Emmanuel C. Gorospe, and Kenneth K. Wang. 2013. “Endoscopic Therapy for Barrett’s Esophagus and Early Esophageal Adenocarcinoma.” *Gastroenterology clinics of North America* 42(1):175–85.

Lei, Tim C., David A. Ammar, Omid Masihzadeh, Emily A. Gibson, and Malik Y. Kahook. 2011. “Label-Free Imaging of Trabecular Meshwork Cells Using Coherent Anti-Stokes Raman Scattering (CARS) Microscopy.” *Molecular vision* 17:2628–33.

Levine, D. S., P. L. Blount, R. E. Rudolph, and B. J. Reid. 2000. “Safety of a Systematic Endoscopic Biopsy Protocol in Patients with Barrett’s Esophagus.” *The American journal of gastroenterology* 95(5):1152–57.

Li, Qingbo, Wei Wang, Xiaofeng Ling, and Jin Guang Wu. 2013. “Detection of Gastric Cancer with Fourier Transform Infrared Spectroscopy and Support Vector Machine Classification.” *BioMed research international* 2013:942427.

Liao, Linda M. et al. 2012. “Nonsteroidal Anti-Inflammatory Drug Use Reduces Risk of Adenocarcinomas of the Esophagus and Esophagogastric Junction in a Pooled Analysis.” *Gastroenterology* 142(3):442–52.e5; quiz e22–23.

Lieber, Chad A., and Anita Mahadevan-Jansen. 2003. “Automated Method for Subtraction of Fluorescence from Biological Raman Spectra.” *Applied spectroscopy* 57(11):1363–67.

Lightdale, Charles J. 2013. “Optical Coherence Tomography in Barrett’s Esophagus.” *Gastrointestinal endoscopy clinics of North America* 23(3):549–63.

Lin, Otto S. et al. 2007. “Blinded Comparison of Esophageal Capsule Endoscopy versus Conventional Endoscopy for a Diagnosis of Barrett’s Esophagus in Patients with Chronic Gastroesophageal Reflux.” *Gastrointestinal endoscopy* 65(4):577–83.

Liu, Lixia et al. 2005. “Significance of the Depth of Tumor Invasion and Lymph Node Metastasis in Superficially Invasive (T1) Esophageal Adenocarcinoma.” *The American journal of surgical pathology* 29(8):1079–85.

Liu, Weitian, Hejin Hahn, Robert D. Odze, and Raj K. Goyal. 2009. “Metaplastic Esophageal Columnar Epithelium without Goblet Cells Shows DNA Content Abnormalities Similar to Goblet Cell-Containing Epithelium.” *The American journal of gastroenterology* 104(4):816–24.

Liu, Xinxue et al. 2014. “Gastro-Esophageal Reflux Disease Symptoms and Demographic Factors as a Pre-Screening Tool for Barrett’s Esophagus.” *PloS one* 9(4):e94163.

Lloyd, Gavin Rhys et al. 2012. “Histological Imaging of a Human Colon Polyp Sample Using Raman Spectroscopy and Self Organising Maps.” *Vibrational Spectroscopy* 60:43–49.

Lloyd, Gavin Rhys et al. 2013. “Discrimination between Benign, Primary and Secondary Malignancies in Lymph Nodes from the Head and Neck Utilising Raman Spectroscopy and Multivariate Analysis.” *The Analyst* 138(14):3900–3908.

Lomo, Leslie C. et al. 2006. “Crypt Dysplasia with Surface Maturation: A Clinical, Pathologic, and Molecular Study of a Barrett’s Esophagus Cohort.” *The American journal of surgical pathology* 30(4):423–35.

Lovat, L. B. et al. 2006. “Elastic Scattering Spectroscopy Accurately Detects High Grade Dysplasia and Cancer in Barrett’s Oesophagus.” *Gut* 55(8):1078–83.

Lu, Shaoying, and Thomas D. Wang. 2008. “In Vivo Cancer Biomarkers of Esophageal Neoplasia.” *Cancer biomarkers : section A of Disease markers* 4(6):341–50.

Lyng, F. M. et al. 2007. “Vibrational Spectroscopy for Cervical Cancer Pathology, from Biochemical Analysis to Diagnostic Tool.” *Experimental and molecular pathology* 82(2):121–29.

Mackanos, Mark A, and Christopher H. Contag. 2010. “Fiber-Optic Probes Enable Cancer Detection with FTIR Spectroscopy.” *Trends in biotechnology* 28(6):317–23.

Mannath, J., V. Subramanian, C. J. Hawkey, and K. Ragunath. 2010. “Narrow Band Imaging for Characterization of High Grade Dysplasia and Specialized Intestinal Metaplasia in Barrett’s Esophagus: A Meta-Analysis.” *Endoscopy* 42(5):351–59.

Martinho, Herculano da Silva et al. 2008. “Role of Cervicitis in the Raman-Based Optical Diagnosis of Cervical Intraepithelial Neoplasia.” *Journal of biomedical optics* 13(5):054029.

Matousek, P. et al. 2005. “Subsurface Probing in Diffusely Scattering Media Using Spatially Offset Raman Spectroscopy.” *Applied Spectroscopy* 59(4):393–400.

Matousek, P., and N. Stone. 2009. “Emerging Concepts in Deep Raman Spectroscopy of Biological Tissue.” *The Analyst* 134(6):1058–66.

Matousek, Pavel. 2009. “Deep, Noninvasive Raman Spectroscopy of Diffuesly Scattering Media.” Pp. 405–26 in *Infrared and Raman Spetroscopic Imaging*, edited by Reiner Salzer and Heinz Siesler. Weinheim: Wiley-VCH.

Matousek, Pavel, and Nicholas Stone. 2013. “Recent Advances in the Development of Raman Spectroscopy for Deep Non-Invasive Medical Diagnosis.” *Journal of biophotonics* 6(1):7–19.

McDonald, Stuart A. C., Trevor A. Graham, Danielle L. Lavery, Nicholas A. Wright, and Marnix Jansen. 2015. “The Barrett’s Gland in Phenotype Space.” *CMGH Cellular and Molecular Gastroenterology and Hepatology* 1(1):41–54.

McDonald, Stuart A. C., Danielle Lavery, Nicholas A. Wright, and Marnix Jansen. 2015. “Barrett Oesophagus: Lessons on Its Origins from the Lesion Itself.” *Nature reviews. Gastroenterology & hepatology* 12(1):50–60.

Menzies, Georgina E. et al. 2013. “Fourier Transform Infrared for Noninvasive Optical Diagnosis of Oral, Oropharyngeal, and Laryngeal Cancer.” *Translational research : the journal of laboratory and clinical medicine*.

Miljković, Miloš, Benjamin Bird, and Max Diem. 2012. “Line Shape Distortion Effects in Infrared Spectroscopy.” *The Analyst* 137(17):3954–64.

Miller, C. T. et al. 2006. “Genomic Amplification of MET with Boundaries within Fragile Site FRA7G and Upregulation of MET Pathways in Esophageal Adenocarcinoma.” Oncogene 25(September 2005):409–18.

Miller, JN and Miller JC. 2010. “Multivariate Analysis.” Pp. 221–50 in *Statistics and chemometrics for anlaytical chemistry*. Harlow, England: Pearson.

Molloy, R. M., and A Sonnenberg. 1997. “Relation between Gastric Cancer and Previous Peptic Ulcer Disease.” *Gut* 40(2):247–52.

Montgomery, E. et al. 2001. “Reproducibility of the Diagnosis of Dysplasia in Barrett Esophagus: A Reaffirmation.” *Human pathology* 32(4):368–78.

Murray, L. et al. 2006. “TP53 and Progression from Barrett’s Metaplasia to Oesophageal Adenocarcinoma in a UK Population Cohort.” *Gut* 55(10):1390–97.

Nallala, Jayakrupakar et al. 2012. “Infrared Spectral Imaging as a Novel Approach for Histopathological Recognition in Colon Cancer Diagnosis.” *Journal of biomedical optics* 17(11):116013.

Nallala, Jayakrupakar, Gavin Rhys Lloyd, and Nicholas Stone. 2015. “Evaluation of Different Tissue de-Paraffinization Procedures for Infrared Spectral Imaging.” *The Analyst* 140(7):2369–75.

Nanda, K. et al. 2000. “Accuracy of the Papanicolaou Test in Screening for and Follow-up of Cervical Cytologic Abnormalities: A Systematic Review.” *Annals of internal medicine* 132(10):810–19.

National Institute for Health and Care Excellence. 2014. Endoscopic Radiofrequency Ablation for Barrett ’s Oesophagus with Low-Grade Dysplasia or No Dysplasia. *NICE*.

National Screening Committee. 2009. “Programme Appraisal Criteria, Criteria for Appraising the Viability, Effectiveness and Appropriateness of a Screening Programme.” *Public Health England*.

Nguyen, Dang M. et al. 2009. “Medication Usage and the Risk of Neoplasia in Patients with Barrett’s Esophagus.” *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 7(12):1299–1304.

Nguyen-Tang, Thai, Jean-Louis Frossard, and Jean-Marc Dumonceau. 2011. “[Barrett’s Esophagus: New Screening Methods and New Endoscopic Therapy].” *Revue médicale suisse* 7(307):1672–77.

Nicholson, Anna M. et al. 2012. “Barrett’s Metaplasia Glands Are Clonal, Contain Multiple Stem Cells and Share a Common Squamous Progenitor.” *Gut* 61(10):1380–89.

Nijssen, Annieke et al. 2007. “Discriminating Basal Cell Carcinoma from Perilesional Skin Using High Wave-Number Raman Spectroscopy.” *Journal of biomedical optics* 12(3):034004.

Odze, R. D. 2006. “Diagnosis and Grading of Dysplasia in Barrett’s Oesophagus.” *Journal of clinical pathology* 59(10):1029–38.

Old, OJ et al. 2014. “Vibrational Spectroscopy for Cancer Diagnostics.” *Analytical methods* 6(12):3901–17.

Olympus medical. 2013. “Olympus GIF-Q40.” Retrieved 20 November 2013 (http://www.olympus.co.uk/medical/en/medical_systems/products_services/product_details/product_details_9089.jsp).

Orman, Eric S., Nan Li, and Nicholas J. Shaheen. 2013. “Efficacy and Durability of Radiofrequency Ablation for Barrett’s Esophagus: Systematic Review and Meta-Analysis.” *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 11(10):1245–55.

Papamarkakis, Kostas et al. 2010. “Cytopathology by Optical Methods: Spectral Cytopathology of the Oral Mucosa.” *Laboratory investigation; a journal of technical methods and pathology* 90(4):589–98.

Patil, Chetan A et al. 2011. “Integrated System for Combined Raman Spectroscopy-Spectral Domain Optical Coherence Tomography.” *Journal of biomedical optics* 16(1):011007.

Pech, O. et al. 2008. “Long-Term Results and Risk Factor Analysis for Recurrence after Curative Endoscopic Therapy in 349 Patients with High-Grade Intraepithelial Neoplasia and Mucosal Adenocarcinoma in Barrett’s Oesophagus.” *Gut* 57(9):1200–1206.

Pech, Oliver et al. 2014. “Long-Term Efficacy and Safety of Endoscopic Resection for Patients With Mucosal Adenocarcinoma of the Esophagus.” *Gastroenterology* 146(3):652–60.e1.

Pera, Manuel, Carlos Manterola, Oscar Vidal, and Luis Grande. 2005. “Epidemiology of Esophageal Adenocarcinoma.” *Journal of surgical oncology* 92(3):151–59.

Peters, J. H. et al. 1994. “Outcome of Adenocarcinoma Arising in Barrett’s Esophagus in Endoscopically Surveyed and Nonsurveyed Patients.” *The Journal of thoracic and cardiovascular surgery* 108(5):813–21; discussion 821–22.

Phoa, K. Nadine et al. 2014. “Radiofrequency Ablation vs Endoscopic Surveillance for Patients with Barrett Esophagus and Low-Grade Dysplasia: A Randomized Clinical Trial.” *JAMA : the journal of the American Medical Association* 311(12):1209–17.

Picardo, Sarah L., Stephen G. Maher, Jacintha N. O’Sullivan, and John V. Reynolds. 2012. “Barrett’s to Oesophageal Cancer Sequence: A Model of Inflammatory-Driven Upper Gastrointestinal Cancer.” Digestive Surgery 29(3):251–60.

Pijanka, Jacek K. et al. 2013. “Identification of Different Subsets of Lung Cells Using Raman Microspectroscopy and Whole Cell Nucleus Isolation.” *The Analyst* 138(17):5052–58.

Pohl, Heiko, Brenda Sirovich, and H. Gilbert Welch. 2010. “Esophageal Adenocarcinoma Incidence: Are We Reaching the Peak?” *Cancer epidemiology, biomarkers & prevention* 19(6):1468–70.

Pouw, Roos E. et al. 2010. “Stepwise Radical Endoscopic Resection for Eradication of Barrett’s Oesophagus with Early Neoplasia in a Cohort of 169 Patients.” *Gut* 59(9):1169–77.

Powell, J., and C. C. McConkey. 1990. “Increasing Incidence of Adenocarcinoma of the Gastric Cardia and Adjacent Sites.” *British journal of cancer* 62(3):440–43.

Powell, Jean, Chris C. McConkey, E. Walford Gillison, and Robert T. Spychal. 2002. “Continuing Rising Trend in Oesophageal Adenocarcinoma.” *International journal of cancer.* 102(4):422–27.

Prasad, Ganapathy A. et al. 2009. “Endoscopic and Surgical Treatment of Mucosal (T1a) Esophageal Adenocarcinoma in Barrett’s Esophagus.” *Gastroenterology* 137(3):815–23.

Qiu, Le et al. 2010. “Multispectral Scanning during Endoscopy Guides Biopsy of Dysplasia in Barrett’s Esophagus.” *Nature medicine* 16(5):603–6, 1p following 606.

Quaroni, Luca, and Alan G. Casson. 2009. “Characterization of Barrett Esophagus and Esophageal Adenocarcinoma by Fourier-Transform Infrared Microscopy.” *The Analyst* 134(6):1240–46.

Quaroni, Luca, Ronghua Zhao, and Alan G. Casson. 2009. “Shining Light on Barrett’s Esophagus.” *Expert review of gastroenterology & hepatology* 3(6):577–80.

Quaroni, Luca, and Theodora Zlateva. 2011. “Infrared Spectromicroscopy of Biochemistry in Functional Single Cells.” *The Analyst* 136(16):3219–32.

Raimes, Simon. 2009. *Oesophagogastric Surgery. A Companion to Specialist Surgical Practice*. Fourth ed. Edited by S Michael Griffin and Simon A Raimes. Saunders Elsevier.

Ramirez, Francisco C., Rodney Akins, and Masud Shaukat. 2008. “Screening of Barrett’s Esophagus with String-Capsule Endoscopy: A Prospective Blinded Study of 100 Consecutive Patients Using Histology as the Criterion Standard.” *Gastrointestinal endoscopy* 68(1):25–31.

Reid, B. J. et al. 1988. “Observer Variation in the Diagnosis of Dysplasia in Barrett’s Esophagus.” *Human pathology* 19(2):166–78.

Reid, B. J., R. C. Haggitt, C. E. Rubin, and P. S. Rabinovitch. 1987. “Barrett’s Esophagus. Correlation between Flow Cytometry and Histology in Detection of Patients at Risk for Adenocarcinoma.” *Gastroenterology* 93(1):1–11.

Rex, Douglas K. et al. 2003. “Screening for Barrett’s Esophagus in Colonoscopy Patients with and without Heartburn.” *Gastroenterology* 125(6):1670–77.

Riddell, Robert H., and Robert D. Odze. 2009. “Definition of Barrett’s Esophagus: Time for a Rethink--Is Intestinal Metaplasia Dead?” *The American journal of gastroenterology* 104(10):2588–94.

Robichaux-Viehoever, Amy et al. 2007. “Characterization of Raman Spectra Measured in Vivo for the Detection of Cervical Dysplasia.” *Applied spectroscopy* 61(9):986–93.

Romagnoli, S. et al. 2001. “Molecular Alterations of Barrett’s Esophagus on Microdissected Endoscopic Biopsies.” *Laboratory investigation; a journal of technical methods and pathology* 81(2):241–47.

Romeo, Melissa, Brian Mohlenhoff, Michael Jennings, and Max Diem. 2006. “Infrared Micro-Spectroscopic Studies of Epithelial Cells.” *Biochimica et biophysica acta* 1758(7):915–22.

Ronkainen, Jukka et al. 2005. “Prevalence of Barrett’s Esophagus in the General Population: An Endoscopic Study.” *Gastroenterology* 129(6):1825–31.

Ross-Innes, Caryn S. et al. 2015. “Evaluation of a Minimally Invasive Cell Sampling Device Coupled with Assessment of Trefoil Factor 3 Expression for Diagnosing Barrett’s Esophagus: A Multi-Center Case–Control Study” edited by Eduardo L. Franco. *PLOS Medicine* 12(1):e1001780.

Rothwell, Peter M. et al. 2011. “Effect of Daily Aspirin on Long-Term Risk of Death due to Cancer: Analysis of Individual Patient Data from Randomised Trials.” *Lancet* 377(9759):31–41.

Rubenstein, Joel H., Amnon Sonnenberg, Jennifer Davis, Laurence McMahon, and John M. Inadomi. 2008. “Effect of a Prior Endoscopy on Outcomes of Esophageal Adenocarcinoma among United States Veterans.” *Gastrointestinal endoscopy* 68(5):849–55.

Rubin, C. E. 1960. “Exfoliative Cytology of the Esophagus.” *CA: a cancer journal for clinicians* 10:90–96.

Saadi, Amel et al. 2010. “Stromal Genes Discriminate Preinvasive from Invasive Disease, Predict Outcome, and Highlight Inflammatory Pathways in Digestive Cancers.” Proceedings of the National Academy of Sciences of the United States of America 107(5):2177–82.

Salmela, M. T., M. L. Karjalainen-Lindsberg, P. Puolakkainen, and U. Saarialho-Kere. 2001. “Upregulation and Differential Expression of Matrilysin (MMP-7) and Metalloelastase (MMP-12) and Their Inhibitors TIMP-1 and TIMP-3 in Barrett’s Oesophageal Adenocarcinoma.” British journal of cancer 85(3):383–92.

Van Sandick, J. W. et al. 1998. “Impact of Endoscopic Biopsy Surveillance of Barrett’s Oesophagus on Pathological Stage and Clinical Outcome of Barrett's Carcinoma.” *Gut* 43(2):216–22.

Sattlecker, Martina, Conrad Bessant, Jennifer Smith, and Nick Stone. 2010. “Investigation of Support Vector Machines and Raman Spectroscopy for Lymph Node Diagnostics.” *The Analyst* 135(5):895–901.

Saxena, Payal, and Marcia Irene Canto. 2013. “Red Flag Imaging Techniques in Barrett’s Esophagus.” *Gastrointestinal endoscopy clinics of North America* 23(3):535–47.

Schlemper, R. J. et al. 2000. “The Vienna Classification of Gastrointestinal Epithelial Neoplasia.” *Gut* 47(2):251–55.

Schubert, Jennifer M. et al. 2010. “Spectral Cytopathology of Cervical Samples: Detecting Cellular Abnormalities in Cytologically Normal Cells.” *Laboratory investigation; a journal of technical methods and pathology* 90(7):1068–77.

Sepesi, Boris et al. 2010. “Are Endoscopic Therapies Appropriate for Superficial Submucosal Esophageal Adenocarcinoma? An Analysis of Esophagectomy Specimens.” *Journal of the American College of Surgeons* 210(4):418–27.

Shaheen, Nicholas et al. 2011. “Durability of Radiofrequency Ablation in Barrett’s Esophagus with Dysplasia.” *Gastroenterology* 141(2):460–68.

Shaheen, Nicholas J. et al. 2009. “Radiofrequency Ablation in Barrett’s Esophagus with Dysplasia.” *The New England journal of medicine* 360(22):2277–88.

Shaheen, Nicholas, and David F. Ransohoff. 2002. “Gastroesophageal Reflux, Barrett Esophagus, and Esophageal Cancer: Scientific Review.” *JAMA : the journal of the American Medical Association* 287(15):1972–81.

Shariff, M. Kareem et al. 2012. “Randomized Crossover Study Comparing Efficacy of Transnasal Endoscopy with that of Standard Endoscopy to Detect Barrett’s Esophagus.” *Gastrointestinal endoscopy* 75(5):954–61.

Sharma, Prateek, Gary W. Falk, et al. 2006. “Dysplasia and Cancer in a Large Multicenter Cohort of Patients with Barrett’s Esophagus.” *Clinical gastroenterology and hepatology* 4(5):566–72.

Sharma, Prateek, John Dent, et al. 2006. “The Development and Validation of an Endoscopic Grading System for Barrett’s Esophagus: The Prague C & M Criteria.” *Gastroenterology* 131(5):1392–99.

Sheng, Daping et al. 2013. “Comparison of Serum from Gastric Cancer Patients and from Healthy Persons Using FTIR Spectroscopy.” *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* 116:365–69.

Shepherd, Neil. 2013. “Barrett’s Oesophagus.” Pp. 39–49 in *Morson and Dawson’s Gastrointestinal Pathology*, edited by Neil A Shepherd et al. Chichester: Wiley-Blackwell.

Shetty, G., C. Kendall, N. Shepherd, N. Stone, and H. Barr. 2006. “Raman Spectroscopy: Elucidation of Biochemical Changes in Carcinogenesis of Oesophagus.” *British journal of cancer* 94(10):1460–64.

Shim, M. G., L. M. Song, N. E. Marcon, and B. C. Wilson. 2000. “In Vivo near-Infrared Raman Spectroscopy: Demonstration of Feasibility during Clinical Gastrointestinal Endoscopy.” *Photochemistry and photobiology* 72(1):146–50.

Short, Michael A. et al. 2008. “Development and Preliminary Results of an Endoscopic Raman Probe for Potential in Vivo Diagnosis of Lung Cancers.” *Optics letters* 33(7):711–13.

Shu, Y. J. 1983. “Cytopathology of the Esophagus. An Overview of Esophageal Cytopathology in China.” *Acta cytologica* 27(1):7–16.

Silva, Francisco Baldaque et al. 2011. “Endoscopic Assessment and Grading of Barrett’s Esophagus Using Magnification Endoscopy and Narrow-Band Imaging: Accuracy and Interobserver Agreement of Different Classification Systems (with Videos).” *Gastrointestinal endoscopy* 73(1):7–14.

Singh, M. et al. 2011. “Observer Agreement in the Assessment of Narrowband Imaging System Surface Patterns in Barrett’s Esophagus: A Multicenter Study.” *Endoscopy* 43(9):745–51.

Singh, S. P., Atul Deshmukh, Pankaj Chaturvedi, and C. Murali Krishna. 2012. “Raman Spectroscopy in Head and Neck Cancers: Toward Oncological Applications.” *Journal of cancer research and therapeutics* 8 Suppl 1:S126–32.

Skacel, Marek et al. 2002. “p53 Expression in Low Grade Dysplasia in Barrett’s Esophagus: Correlation with Interobserver Agreement and Disease Progression.” *The American journal of gastroenterology* 97(10):2508–13.

Skinner, D. B. et al. 1983. “Barrett’s Esophagus. Comparison of Benign and Malignant Cases.” *Annals of surgery* 198(4):554–65.

Smith, Kylie J. et al. 2005. “Interactions among Smoking, Obesity, and Symptoms of Acid Reflux in Barrett’s Esophagus.” *Cancer epidemiology, biomarkers & prevention* 14(11 Pt 1):2481–86.

Spechler, Sj. 1997. “Barrett’s Esophagus: Should We Brush off This Ballooning Problem?” *Gastroenterology* 112(6):2138–42.

Spechler, Stuart J., Prateek Sharma, Rhonda F. Souza, John M. Inadomi, and Nicholas J. Shaheen. 2011. “American Gastroenterological Association Medical Position Statement on the Management of Barrett’s Esophagus.” *Gastroenterology* 140(3):1084–91.

Srivastava, Amitabh et al. 2007. “Morphologic Features Are Useful in Distinguishing Barrett Esophagus from Carditis with Intestinal Metaplasia.” *The American journal of surgical pathology* 31(11):1733–41.

Stein, Hubert J. et al. 2005. “Early Esophageal Cancer: Pattern of Lymphatic Spread and Prognostic Factors for Long-Term Survival after Surgical Resection.” *Annals of surgery* 242(4):566–73; discussion 573–75.

Stone, N., P. Stavroulaki, C. Kendall, M. Birchall, and H. Barr. 2000. “Raman Spectroscopy for Early Detection of Laryngeal Malignancy: Preliminary Results.” *The Laryngoscope* 110(10 Pt 1):1756–63.

Stone, Nicholas, Maria Consuelo Hart Prieto, Paul Crow, Jeremy Uff, and Alistair William Ritchie. 2007. “The Use of Raman Spectroscopy to Provide an Estimation of the Gross Biochemistry Associated with Urological Pathologies.” *Analytical and bioanalytical chemistry* 387(5):1657–68.

Stone, Nicholas, Catherine Kendall, and Hugh Barr. 2008. “Applications of Vibrational Spectroscopy in Medical Diagnosis: Raman Spectroscopy as a Potential Tool for Early Diagnosis of Malignancies in Esophageal and Bladder Tissues.” in *Vibrational spectroscopy for medical diagnostics*, edited by M Diem, J Chambers, and P Griffiths. Wiley.

Stone, Nicholas, Catherine Kendall, Jenny Smith, Paul Crow, and Hugh Barr. 2004. “Raman Spectroscopy for Identification of Epithelial Cancers.” *Faraday Discussions* 126(May 2003):141–57.

Streitz, J. M., C. W. Andrews, and F. H. Ellis. 1993. “Endoscopic Surveillance of Barrett’s Esophagus. Does It Help?” *The Journal of thoracic and cardiovascular surgery* 105(3):383–87; discussion 387–88.

Su, Zhan et al. 2012. “Common Variants at the MHC Locus and at Chromosome 16q24.1 Predispose to Barrett’s Esophagus.” *Nature genetics* 44(10):1131–36.

Subramanian, Venkataraman, and Krish Ragunath. 2013. “Advanced Endoscopic Imaging: A Review of Commercially Available Technologies.” *Clinical gastroenterology and hepatology 12(3):368-76*.

Suter, Melissa J. et al. 2009. “Comprehensive Microscopy of the Esophagus in Human Patients With Optical Frequency Domain Imaging.” *Gasrorintestinal Endoscopy* 68(4):745–53.

Takubo, K., J. M. Nixon, and J. R. Jass. 1995. “Ducts of Esophageal Glands Proper and Paneth Cells in Barrett’s Esophagus: Frequency in Biopsy Specimens.” *Pathology* 27(4):315–17.

Taylor, Justin B., and Joel H. Rubenstein. 2010. “Meta-Analyses of the Effect of Symptoms of Gastroesophageal Reflux on the Risk of Barrett’s Esophagus.” *The American journal of gastroenterology* 105(8):1729, 1730–37; quiz 1738.

Tfaili, Sana et al. 2012. “Confocal Raman Microspectroscopy for Skin Characterization: A Comparative Study between Human Skin and Pig Skin.” *The Analyst* 137(16):3673–82.

Thekkek, Nadhi et al. 2013. “Modular Video Endoscopy for in Vivo Cross-Polarized and Vital-Dye Fluorescence Imaging of Barrett’s-Associated Neoplasia.” *Journal of biomedical optics* 18(2):26007.

Tomizawa, Yutaka et al. 2013. “Assessment of the Diagnostic Performance and Interobserver Variability of Endocytoscopy in Barrett’s Esophagus: A Pilot Ex-Vivo Study.” *World journal of gastroenterology : WJG* 19(46):8652–58.

Toms, Steven A., Peter E. Konrad, Wei-Chiang Lin, and Robert J. Weil. 2006. “Neuro-Oncological Applications of Optical Spectroscopy.” *Technology in cancer research & treatment* 5(3):231–38.

Townsend, Douglas et al. 2015. “Infrared Micro-Spectroscopy for Cyto-Pathological Classification of Esophageal Cells.” *The Analyst* 140(7):2215–23.

Trivers, Katrina F., Susan a Sabatino, and Sherri L. Stewart. 2008. “Trends in Esophageal Cancer Incidence by Histology, United States, 1998-2003.” *International journal of cancer.* 123(6):1422–28.

Tu, Haohua, and Stephen a Boppart. 2014. “Coherent Anti-Stokes Raman Scattering Microscopy: Overcoming Technical Barriers for Clinical Translation.” *Journal of biophotonics* 7(1-2):9–22.

Uckermann, Ortrud et al. 2014. “Label-Free Delineation of Brain Tumors by Coherent Anti-Stokes Raman Scattering Microscopy in an Orthotopic Mouse Model and Human Glioblastoma.” *PloS one* 9(9):e107115.

Urquhart, P., R. DaCosta, and N. Marcon. 2013. “Endoscopic Mucosal Imaging of Gastrointestinal Neoplasia in 2013.” *Current gastroenterology reports* 15(7):330.

Vaezi, M. F., and J. E. Richter. 1996. “Role of Acid and Duodenogastroesophageal Reflux in Gastroesophageal Reflux Disease.” *Gastroenterology* 111(5):1192–99.

Vaezi, Michael F., and Peter J. Kahrilas. 2013. “Barrett’s Esophagus Surveillance: Time to Rethink If One Size Fits All?” *Gastroenterology* 145(3):503–5.

Varadarajulu, Shyam et al. 2011. “GI Endoscopes.” *Gastrointestinal endoscopy* 74(1):1–6.e6.

Wang, Jiansheng et al. 2010. “Evaluation of Gallbladder Lipid Level during Carcinogenesis by an Infrared Spectroscopic Method.” *Digestive Diseases and Sciences* 55(9):2670–75.

Wang, Jian-sheng et al. 2003. “FT-IR Spectroscopic Analysis of Normal and Cancerous Tissues of Esophagus.” 9(9):1897–99.

Wang, Kenneth K., and Richard E. Sampliner. 2008. “Updated Guidelines 2008 for the Diagnosis, Surveillance and Therapy of Barrett’s Esophagus.” *The American journal of gastroenterology* 103(3):788–97.

Wang, Thomas D. et al. 2007. “Detection of Endogenous Biomolecules in Barrett’s Esophagus by Fourier Transform Infrared Spectroscopy.” *Proceedings of the National Academy of Sciences of the United States of America* 104(40):15864–69.

Wani, Sachin et al. 2011. “Risk Factors for Progression of Low-Grade Dysplasia in Patients with Barrett’s Esophagus.” *Gastroenterology* 141(4):1179–86, 1186.e1.

Ward, Eric M. et al. 2006. “Barrett’s Esophagus Is Common in Older Men and Women Undergoing Screening Colonoscopy regardless of Reflux Symptoms.” *The American journal of gastroenterology* 101(1):12–17.

Watson, A., RC Heading, and Neil A Shepherd. 2005. “Guidelines for the Diagnosis and Management of Barrett ’s Columnar-lined Oesophagus.” *British Society of Gastroenterology.* Loughborough.

Wax, Adam, Neil G. Terry, Evan S. Dellon, and Nicholas J. Shaheen. 2011. “Angle-Resolved Low Coherence Interferometry for Detection of Dysplasia in Barrett’s Esophagus.” *Gastroenterology* 141(2):443–47, 447.e1–2.

Van de Weert, M., J. Hering, and P. Haris. 2005. *Methods for Structural Analysis of Protein Pharmaceuticals*. edited by W Jiskoot and D Crommelin.

Westerterp, Marinke et al. 2005. “Outcome of Surgical Treatment for Early Adenocarcinoma of the Esophagus or Gastro-Esophageal Junction.” *Virchows Archiv : an international journal of pathology* 446(5):497–504.

Wilson, JMG, and J. Jungner. 1968. *Principles and Practice of Screening for Disease*. Geneva.

Wolfsen, Herbert C. et al. 2008. “Prospective, Controlled Tandem Endoscopy Study of Narrow Band Imaging for Dysplasia Detection in Barrett’s Esophagus.” *Gastroenterology* 135(1):24–31.

Wong Kee Song, Louis-Michel et al. 2005. “Diagnostic Potential of Raman Spectroscopy in Barrett’s Esophagus.” Pp. 140–46 in *Proceedings of SPIE*, edited by Tuan Vo-Dinh, Warren S. Grundfest, David A. Benaron, and Gerald E. Cohn.

Wong Kee Song, Louis-Michel, and Brian C. Wilson. 2005. “Endoscopic Detection of Early Upper GI Cancers.” *Best practice & research. Clinical gastroenterology* 19(6):833–56.

Wong, P. T., Rita K. Wong, Thomas A. Caputo, Thomas A. Godwin, and Basil Rigas. 1991. “Infrared Spectroscopy of Exfoliated Human Cervical Cells: Evidence of Extensive Structural Changes during Carcinogenesis.” *Proceedings of the National Academy of Sciences of the United States of America* 88(24):10988–92.

Wood, B. R. et al. 2004. “Fourier Transform Infrared (FTIR) Spectral Mapping of the Cervical Transformation Zone, and Dysplastic Squamous Epithelium.” *Gynecologic oncology* 93(1):59–68.

Yang, P. C., and S. Davis. 1988. “Incidence of Cancer of the Esophagus in the US by Histologic Type.” *Cancer* 61(3):612–17.

Yu, Guo, Guojin Zhang, Carol R. Flach, and Richard Mendelsohn. 2013. “Vibrational Spectroscopy and Microscopic Imaging: Novel Approaches for Comparing Barrier Physical Properties in Native and Human Skin Equivalents.” *Journal of biomedical optics* 18(6):061207.

Zagari, R. M. et al. 2008. “Gastro-Oesophageal Reflux Symptoms, Oesophagitis and Barrett’s Oesophagus in the General Population: The Loiano-Monghidoro Study.” *Gut* 57(10):1354–59.

Zou, Duowu et al. 2011. “Epidemiology of Symptom-Defined Gastroesophageal Reflux Disease and Reflux Esophagitis: The Systematic Investigation of Gastrointestinal Diseases in China (SILC).” *Scandinavian journal of gastroenterology* 46(2):133–41.