HIGH RESOLUTION INFRARED SPECTROSCOPIC ANALYSIS OF HUMAN COLORECTAL PATHOLOGY

Submitted by Rebecca Katie Louise Griggs to the University of Exeter

As a thesis for the degree of

Masters by Research in Physics

In 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: …………………………………………………………..

REBECCA KATIE LOUISE GRIGGS

HIGH RESOLUTION INFRARED SPECTROSCOPIC ANALYSIS OF HUMAN COLORECTAL PATHOLOGY

COLLEGE OF ENGINEERING, MATHEMATICS AND PHYSICAL SCIENCES

Masters by Research

(MRes)

Academic year: 2014-2015

Academic Supervisor: Professor N Stone; Dr C Kendall, Dr G Lloyd

Clinical Supervisors: Prof NA Shepherd; Prof H Barr; Mr T Cook

ABSTRACT

Epithelial misplacement is a benign pathology of the intestines that mimics invasive carcinoma and leads to a high degree of diagnostic difficulty in the discrimination between the two by pathologists. Particular difficulties are associated with the diagnosis between the two, and currently a National Expert Board of specialist pathologists delivers a second opinion for equivocal cases referred from local Hospitals. Novel analysis techniques to assist the diagnosis of epithelial misplacement vs. Cancer would not only aid the Board in the diagnosis, but speed up time taken to diagnosis for future patient management, and therefore increase overall patient satisfaction.

This research aims to define the use of high resolution spectroscopy in the colon and assess its use in analysis of formalin-fixed paraffin-embedded tissue that has been routinely processed by the pathology laboratory. It is predicted that it will be an accurate, rapid and non-destructive novel diagnostic technique, and particularly useful in difficult cases such as in the discrimination of epithelial misplacement from cancer, making it a useful addition to the diagnostic process to improve efficiency of the patient diagnostic and treatment pathway.

65 colonic samples were analysed from a total of 58 patients diagnosed with various colonic pathologies and spectra measured in high resolution using the Agilent® 620 FTIR microscope coupled to a 670 Agilent spectrometer in transmission mode. Spectra were then fed into a specialist computer software programme, analysed using principal component-fed linear discriminant analysis model and tested with a leave one sample out cross validation method.

The 'Two Group' Model defined the ability of the high resolution spectral analysis to discriminate between epithelial misplacement and cancer with a sensitivity of 74.6% and specificity of 82.3%. These results are excellent and have never been recorded previously in the literature. However, further research is required to assess its role in the discrimination of various other colonic pathologies, and the model needs independent testing with new cases of both epithelial misplacement and cancer to continue the assessment of the novel high resolution technique.

ACKNOWLEDGEMENTS

CONTENTS

LIST OF ABBREVIATIONS

AFI	Autofluorescence imaging	LGD	Low grade dysplasia
ANN	Artificial neuronal network	LIFS	Laser-induced fluorescence spectroscopy
ANOVA	Analysis of variance	LOSOCV	Leave one sample out cross validation
APC	Antigen-presenting cell		
ATR	Attenuated total reflectance	MDT	Multi-disciplinary team
BCSP	Bowel cancer screening programme	MINERVA	MId- to NEaR infrared spectroscopy for improVed medical diAgnostics
C	Carbon		
Ca	Calcium	MSI	Micro satellite instability
CRC	Colorectal cancer	N	Nitrogen
CaF_2	Calcium fluoride	NA	Nucleic acid
DNA	Deoxyribonucleic acid	NBI	Narrow band imaging
EB	Expert Board	NHS	National health service
EM	Epithelial misplacement	NICE	National Institute for Clinical Excellence
EMR	Endoscopic mucosal resection		
EMSC	Extended multiplicative signal correction	O	Oxygen
		OCT	O computer tomography
ETMI	Endoscopic trimodal imaging	PC	Principal component
FICE	Flexible image colour enhancement	PCA	Principal component analysis
		PE	Perkin Elmer
FOB	Faecal occult blood	PO_{2-}	Phosphate
FPA	Focal Plane Array	RNA	Ribonucleic acid
FTIR	Fourier transform IR	SD	Standard deviation
GI	Gastrointestinal	SMA	Superior mesenteric artery
H	Hydrogen	STK1	Serological thymidine kinase
H&E	Haematoxylin & eosin	TEMS	Transanal endoscopic mucosal resection
HGD	High grade dysplasia		
IEN	Intra epithelial neoplasia	TNM	Tumour node metastasis
IMA	Inferior mesenteric artery	UK	United kingdom
IR	Infrared	3D	3-dimensional
LDA	Linear discriminant analysis		

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX A FIGURES

CHAPTER 1

THE COLORECTUM AND COLORECTAL CANCER

This thesis presents the use of high resolution infrared spectroscopy as a potential diagnostic tool for the detection of cancer in the colon and in particular the differentiation of cancer from other benign pathological phenomena. Chapter 1 introduces the colon and rectum and provides a general background of current UK colorectal cancer management, highlighting the incidence and rationale for population screening along with stage and classification of colorectal cancer. The management of various colorectal polyps is discussed, and a conclusion is reached with a general introduction to the benign pathological phenomenon of epithelial misplacement, which this thesis focuses on again in later chapters.

1.1 The Colon and Rectum

The large intestine, or colon, is a long muscular tube, the primary function of which is for passage of waste material through to the rectum where it is stored a variable amount of time before expulsion, and secondarily for water re absorption along its length. Fairly liquid-type stool passes from the small intestine where the majority of nutrient absorption occurs, into the large intestine and water is reabsorbed along its length to form hard, formed stool at the rectum.

1.1.1 Anatomy

The large intestine compromises the appendix, caecum, colon and rectum and is approximately 150cm long with a diameter of between 3.5cm and 5cm in the adult (see Figure 1.1).
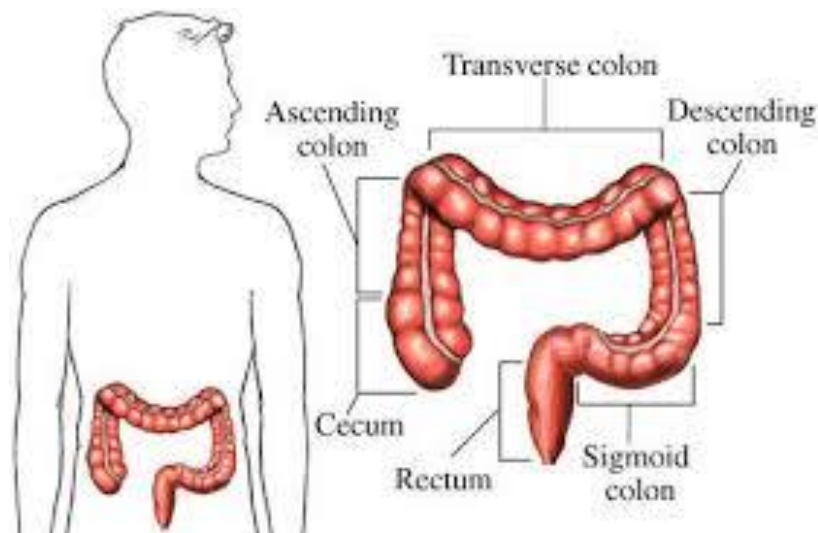
Figure 1.1: Anatomy of the large bowel(Genericlook.com 2010)

Along the length of the colon are markings called haustra which give the colon its distinctive concertina or saccular appearance. They are formed because of bands running the length of the colon called *taeniae coli* which are shorter than the length of the colon and cause the colon to be sacculated between, forming the haustra. Haustral contractions are partly responsible for the movement of waste along the colon and occur approximately every 25 minutes and push the waste from one haustra to the next en mass. In combination with peristaltic movement of the muscle bowel wall, food is moved from the caecum to the rectum in approximately 6 hours, depending on the individual (can take days in others).

1.1.2 Blood supply

The blood supply to the colon is from two main branches off the abdominal aorta, the superior mesenteric artery (SMA) and inferior mesenteric artery (IMA). The SMA gives the middle, right and ileo-colic arteries- supplying caecum to 2/3 of the way along the transverse colon. A marginal artery variably connects SMA branches to branches from the IMA system which then supplies the descending colon, sigmoid colon and recto-sigmoid junction. Lymphatic drainage to local nodes follows the blood supply.

1.1.3 Histology

The large intestine has a similar general histological structure along its length. It is composed throughout of relatively flat mucosa with only a small variation in the amount of lymphoid tissue, lymphoglandular complexes and epithelial constituents.(Shepherd 2002) The normal layers of the bowel wall can be seen in Figure 1.2.
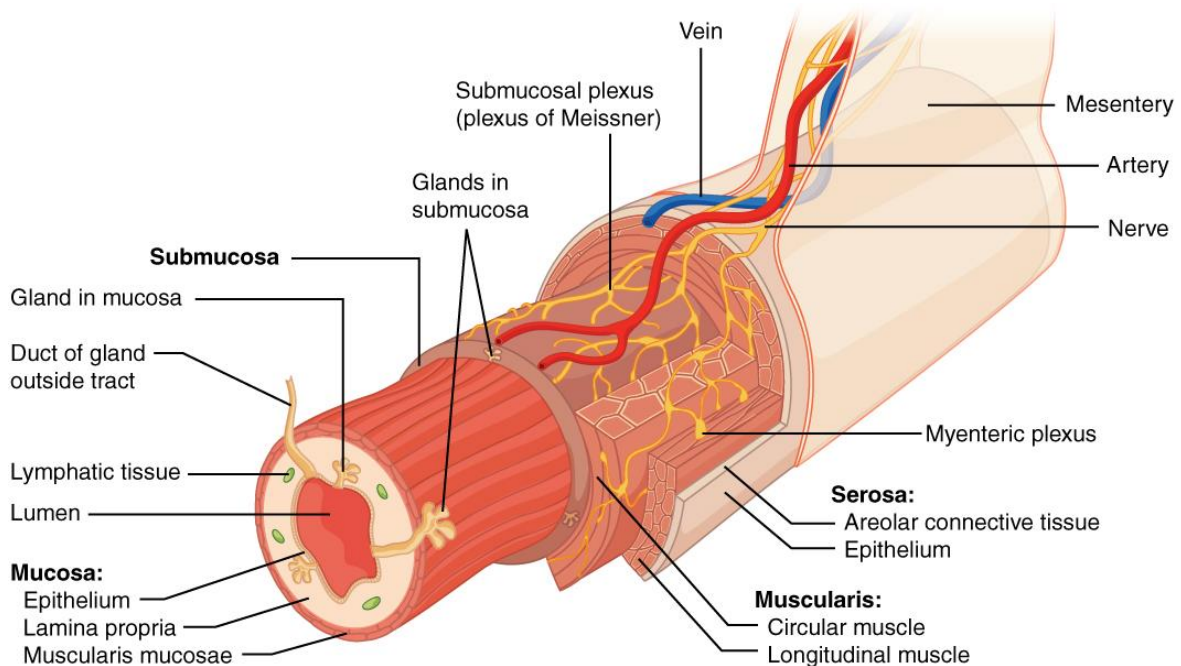
Figure 1.2: Histological cross section of large bowel wall(CNX 2014)

Most of these layers can be seen in the biopsy specimen in Figure 1.3 which had been stained with Haematoxylin and Eosin to give the structures a particular colour. Haematoxylin is a natural dye which stains the nucleus of the cell and cytoplasm rich in ribonucleoprotein a blue/black colour. Eosin is an aniline acid dye that stains cytoplasm, connective tissue and muscle pink/orange. Thus cell structures and tissue can be differentiated according to particular depth of stain and pattern.
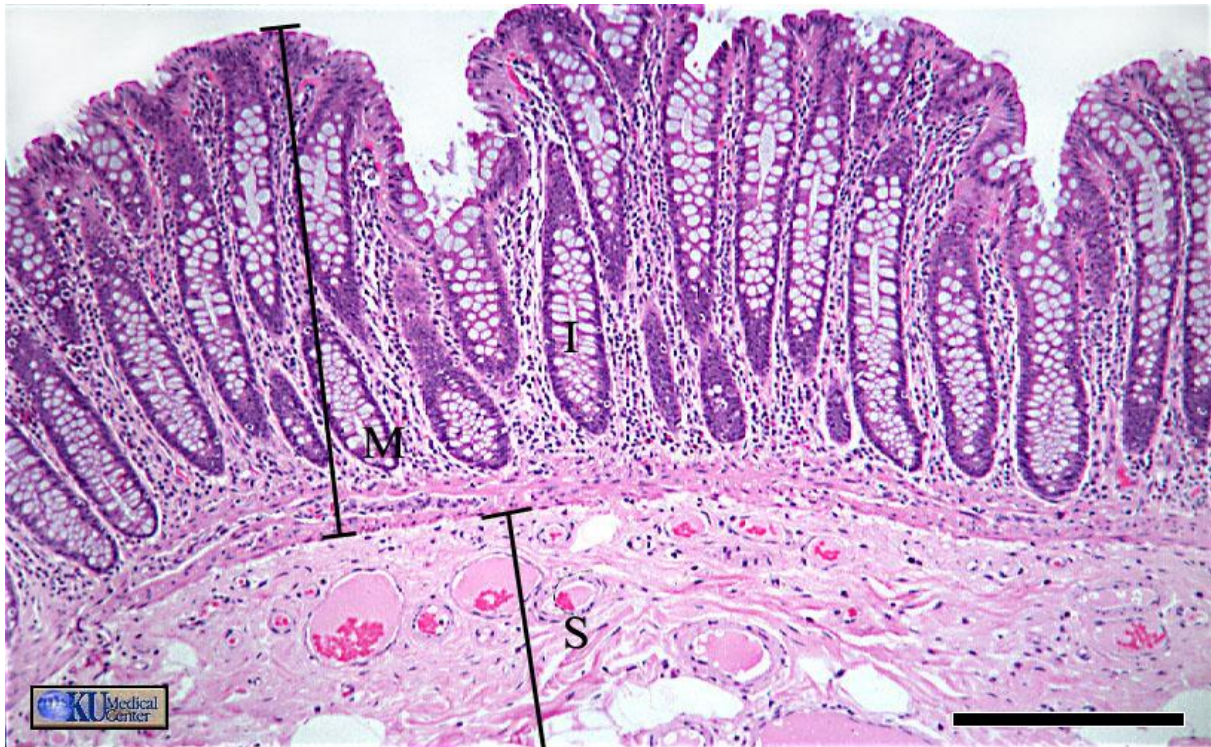
Figure 1.3: Haematoxylin and Eosin stained histological cross section of large bowel mucosa and submucosa(Parsa 2008)

The epithelium is the outermost layer of mucosa and forms the inner bowel wall. It is composed of cell crypts arranged in parallel and lined by cells that secrete mucin, which aids movement of the stool along the length of the gut, and absorptive cells. It is from this layer that adenocarcinomas arise.

The proliferation zone is found at the base of the crypts where most new cell formation occurs and a high volume of dividing cells can be seen here. The crypts are embedded in lamina propria, the connective tissue of the mucosa. The mucosa sits on a submucosal layer of connective tissue which contains vascular, lymphatic and parasympathetic cells (Meissner's submucosal plexus).

A thin muscle layer, the muscularis mucosae, separates the mucosa from the submucosa. Outside the submucosa is the muscularis propria, a thick muscle layer consisting of inner circular and outer longitudinal smooth muscle layers separated by a layer of connective tissue containing Auerbach's myenteric plexus.

The subserosa, another connective tissue layer, and the serosa surround the large intestine, the outermost layer of which, in some areas, is in continuity with the layers of the abdominal wall known as the peritoneum.

1.2 Colorectal cancer in the UK

Bowel cancer is the fourth most common cancer in the UK (2010), accounting for 13% of all new cases. It is the third most common cancer in both males (14% of the male total) and females (11%).(Cancer Research UK 2010)

In 2011, there were 41,581 new cases of bowel cancer in the UK: 56% in men and 44% in women, giving a male: female ratio of 1.3:1. The crude incidence rate shows that there are around 75 new bowel cancer cases for every 100,000 males in the UK and around 56 for every 100,000 females.(Public Health England 2013) Approximately 55% of adult bowel cancer patients (54% of men and 56% of women) in England survived their cancer for five years or more in 2005-2009.

Bowel cancer incidence is strongly related to age, with the highest incidence rates being in older men and women. In the UK between 2009 and 2011, an average 43% of bowel cancer cases were diagnosed in people aged 75 years and over, and 95% were diagnosed in those aged 50 and over, see Figure 1.4.
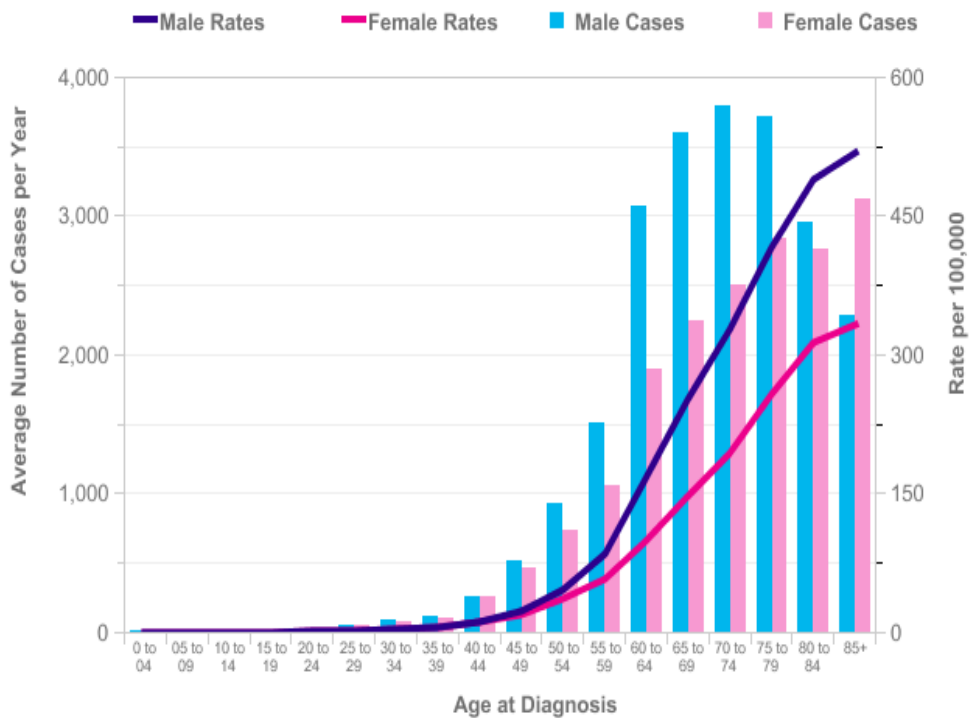
Figure 1.4: The average number of cancer cases per year in the UK(Cancer Research UK 2010)

1.3 National Bowel Cancer Screening

There is a time delay in the development of bowel cancer which gives a window of opportunity for effective population screening. Early diagnosis and removal of polyps reduces the incidence of colorectal cancer and earlier detection of colorectal cancer helps to identify these lesions at an earlier stage.(Loughrey and Shepherd 2015) Regular bowel cancer screening has been shown to reduce the risk of dying from bowel cancer by 16%.(Hewitson et al. 2008) One study investigating the use of a one-off flexible sigmoidoscopy found a 31% reduction in mortality in the screened group compared to the control group.(Atkin et al. 2010)

The diagnosis of bowel abnormalities is either via history, examination and special investigation requested by a clinician or can be via the NHS Bowel Cancer Screening Programme (BCSP). Screening is offered to the whole UK population every two years to all men and women aged 60 to 69 (although by September 2012, 40 of the 58 local screening centres had also started inviting

to an extended population of up to 75 years old). A pilot scheme offering flexible sigmoidoscopy in six bowel cancer screening centres has been introduced, offered on a one-off basis between the ages of 55 and 64, to be rolled out nationally by 2016.(Public Health England 2013)

1.3.1 The National Bowel Cancer Screening Programme

Men and women eligible for screenings receive an invitation letter explaining the programme, and information leaflet, "Bowel Cancer Screening - The Facts".

Figure 1.5: Bowel cancer screening home kit(Public Health England 2013)

About a week later, a faecal occult blood (FOB) test kit is sent out along with step-by-step instructions for completing the test at home and sending the samples to the hub laboratory (Figure 1.5). This test is specifically designed to detect a tiny volume of blood in the faeces at an early stage before the person can see it themselves.

The test is then processed and the results sent within two weeks. 94% of the tested population will receive a **normal** result and will be returned to routine screening and invited for bowel cancer screening every two years hence (if still within the eligible age range for routine screening) and 2% will receive an **abnormal** result. They will be referred for further investigation and usually offered a colonoscopy. 4% receive an **unclear** result and the FOB test is repeated. Most people who repeat the test will then go on to receive a normal result.(Public Health England 2013)

A colonoscopy is an investigation routinely performed in the outpatient setting. It requires the patient to take a laxative at home for the preceding two days, coupled with a low-residue diet to clear out the bowel and thus give the endoscopist the best possible view of the bowel wall on the day of the procedure, enabling them to identify any abnormal pathology more easily. The patient is admitted to hospital usually only for one day, and the colonoscopy is performed with the patient awake but often sedated. It enables the clinical endoscopist to look directly at the lining of the large bowel. It is an excellent tool to detect overt pathology in the large bowel including inflammation, polyps and suspected cancer. A colonoscope is passed into the back passage and guided around the bowel using a wheel controller connecting the scope to a monitor. The benefit of endoscopy over radiological investigation is that it gives the clinician opportunity to biopsy any abnormality found, and remove polyps via the instrument channel of the endoscope (polypectomy), and if the surgeon performs the endoscopy, for future operation planning if necessary.

Polyps may be excised whole with a wire snare combined with diathermy. All tissue is sent to the histopathology lab for further analysis. Of those patients who undergo a screening colonoscopic investigation;

> ➢ Five in 10 people will have a normal result.
> ➢ Four in 10 will be found to have a polyp
> ➢ One in 10 people will be found to have cancer

All tissue taken at the time of colonoscopy is placed in formaldehyde and then sent to the histopathology laboratory where it is embedded in paraffin. Thin tissue sections, three to four micrometers in thickness, are prepared from the processed paraffin blocks on to a glass slide and sent for routine staining which change colour in the presence or absence of varying pathologies (haematoxylin and eosin, H&E). A consultant Histopathologist then reviews the slides and reports any abnormal pathology at a cellular level. For those patients who are discovered to have a polyp on their colonoscopy there is a specific NICE guideline for managing follow-up (NICE Clinical Guideline 118, 2011).(National Institute for Health and Care Excellence 2011). Patients are stratified according to risk (size number and type of polyp) and plan made for repeat colonoscopic follow-up surveillance at a time depending on the risk assessment.

1.4 Management of Colorectal Cancer

Guidelines exist from the National Institute for Health and Care Excellence (NICE) to standardise the care received by patients with bowel cancer throughout the NHS in the UK. Management of colorectal cancer in secondary care begins with appropriate diagnosis of the disease and this is based on a number of special investigations which are offered to the patient (usually including tissue diagnosis and CT scan to check for disease spread). The stage of the cancer can then be assessed according to the Tumour Node Metastasis (TNM) scoring system(Loughrey et al. 2014) (See Table 1.1). It should be emphasised that, in the UK, pathologists use the TNM 5 modification of the TNM system as subsequent versions are not based on sound literature evidence and have not found favour in the UK and Western Europe. The TNM 5 version is reproduced below.

	TUMOUR		NODE		METASTASIS
T1	tumour confined to submucosa	N0	no lymph nodes contain tumour cells	M0	no metastases to distant organs
T2	tumour into (but not through) the muscularis propria	N1	tumour cells present in up to 3 regional lymph nodes	M1	metastases to distant organs
T3	tumour into (but not through) the subserosa	N2	tumour cells present in 4 or more regional lymph nodes		
T4	Tumour penetrated through the serosa and peritoneal surface. **T4a** = extends directly into other nearby organs/structures **T4b =** with **peritoneal involvement only.**				

Table 1.1: Tumour Node Metastasis (TNM) classification.

Stage of disease can then be ascertained using the TMN score and evidence of spread and graded according to histological type (low-high), see Figure 1.6. Morbidity and mortality outcome figures are predicted according to stage and grade of tumour at presentation.
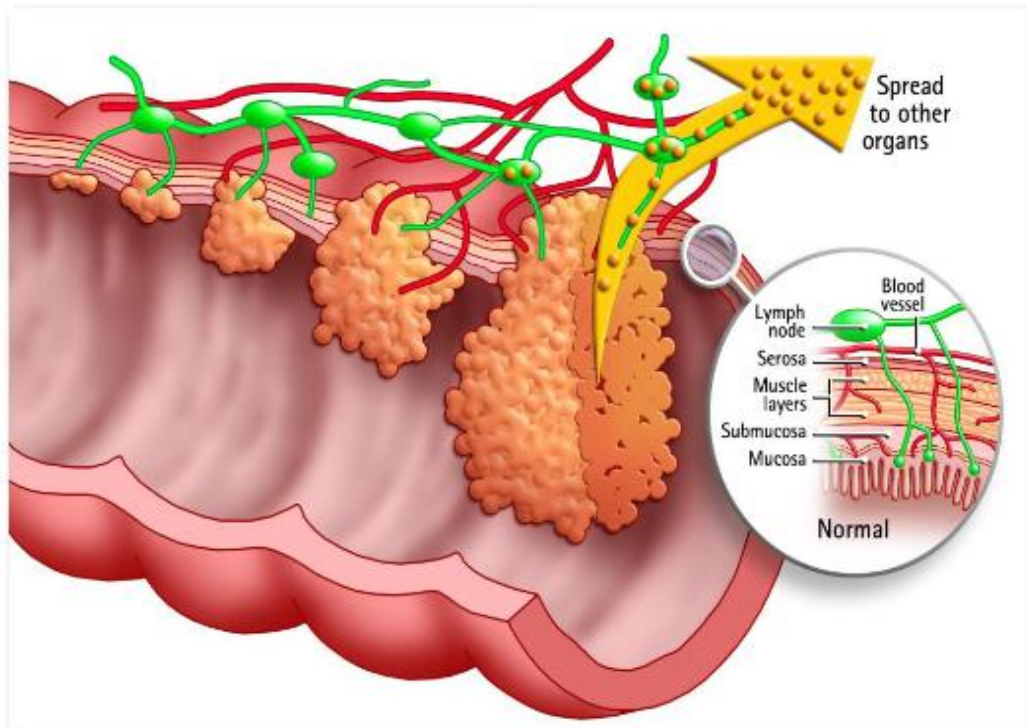
Figure 1.6: A schematic representation of colorectal cancer growth and spread(Fraser 2009)

Practical management of patients varies according to their TMN stage (Clinical Guideline 131). Depending on whether the cancer can be removed by a surgical operation, the patient may be offered pre-operative chemotherapy to try and reduce the bulk of the tumour invasion and enable operative management. The patient may then be offered adjuvant chemotherapy following their operation according to the stage of their disease with the aim of targeting any cancer cells in the blood or lymph circulation and reduce the risk of local or distant recurrence. This whole process has to be done in accordance with the NHS National Service Framework 62 day cancer treatment target (time from referral to treatment).(NHS England 2011)

1.4.1 Management of Early Polyp cancers

Patients found to have a polyp cancer are by definition pT1 as the tumour is confined to the submucosa of the polyp (identification at an early stage means that the cancer has not spread far though the bowel wall, see Figure 1.7 and 1.8, and Table 1.1). The mainstay of treatment for early cancers is local resection. This is usually in the form of endoscopic polypectomy at the time of colonoscopy. A more formal resection may be advocated on the basis of the local resection margin. If the cancer is very close or present at the polypectomy margin, the clinical decision may be for surgical intervention, i.e. colectomy, depending on the outcome of the local colorectal cancer multidisciplinary team meeting discussion after their assessment of the patient's risks and benefits for surgery. Patients fit for surgery will have the advantage of detecting nodal metastasis and a final formal pathological nodal stage that will decide if adjuvant chemotherapy is required.
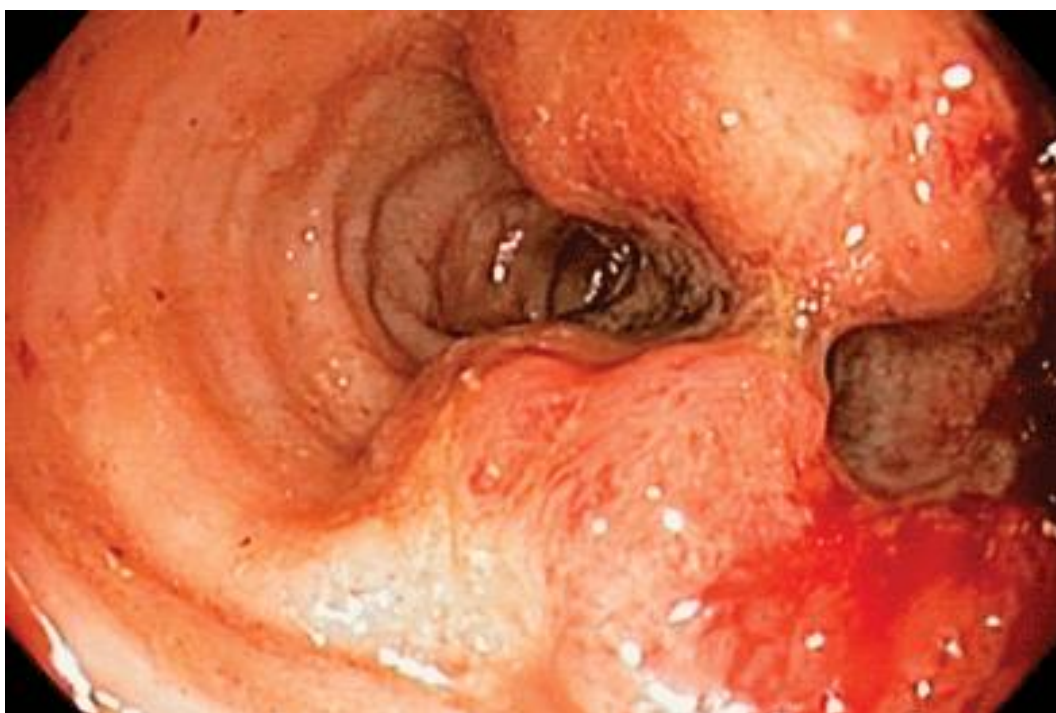
Figure 1.7: A colonoscopic view of an advanced colon cancer with a large ulcerated and bleeding polypoid tumour projecting into the lumen of the bowel(Roy et al. 2012)

Figure 1.8: A colonoscopic view of an early polyp cancer amenable to polypectomy.(Sherman 2009) **The endoscopic appearance is very similar to that of an adenoma at the time of colonoscopy**

1.5 Colorectal cancer

The natural history of colorectal cancer is well documented, and a neoplastic (tumour- forming) precursor identified, the adenoma or adenomatous polyp. The adenoma-carcinoma sequence has been researched in great depth and it is generally accepted that the majority of colorectal cancers arise from such adenomas.(Lockhart-Mummery 1928; Muto et al. 1975; Winawer et al. 1997)

A near obligate precursor to cancer, intra-epithelial neoplasia (IEN) occurs in the epithelial tissues of the mucosa, as low or high grade dysplasia, and shares phenotypic and genotypic similarities with invasive disease on the causal pathway leading from normal tissue to cancer. In addition, IEN serves as a significant risk marker for cancer. Subjects with IEN, particularly those with high grade IEN, are at significantly higher risk than unaffected individuals of developing invasive cancer in the same tissues.(Kelloff et al. 2004) Only a small

proportion of polyps (1-10%) develop into invasive bowel cancer.(Scholefield 2000) Indicators for progression from adenomas to cancer include large size, villous histology and high grade dysplasia.(Terry et al. 2002)

Dysplasia is defined as "an unequivocal neoplastic alteration of epithelium which has the potential to progress to invasive malignancy but remains confined within the basement membrane of the epithelium within which it arose".(Riddell et al. 1983) Dysplasia is classified as either low grade (LGD) or high grade (HGD) based on its histological appearances. HGD has a higher malignant potential than LGD and malignant transformation classically occurs through a stepwise progression of the pathology.

1.5.1 Pathogenesis of Colorectal Cancer

The progression of normal cells to cancer cells is multi-factorial. There has been much debate as to the influence of diet and lifestyle choice, such as cigarette smoking and obesity to the risk of cancer.(Hsing et al. 1998; Terry et al. 2002) It is estimated that 35% of colorectal cancer (CRC) risk may be explained by heritable factors.(Lichtenstein et al. 2000) Inherited genetic factors have an influence in a subset of patients who develop CRC: certain families have hereditary gene mutations which predispose them to developing colorectal cancers (and often at a younger age) such as Lynch syndrome, Gardner's syndrome, effectively a subtype of Lynch syndrome and many more (several of which are uncharacterised).(Sehgal et al. 2014; Zhang et al. 2014; Shiovitz et al. 2014) There is also an established association with some inflammatory bowel conditions and CRC, especially ulcerative colitis and Crohn's disease.(Eaden et al. 2001; Mudter and Neurath 2007; Atreya and Neurath 2008)
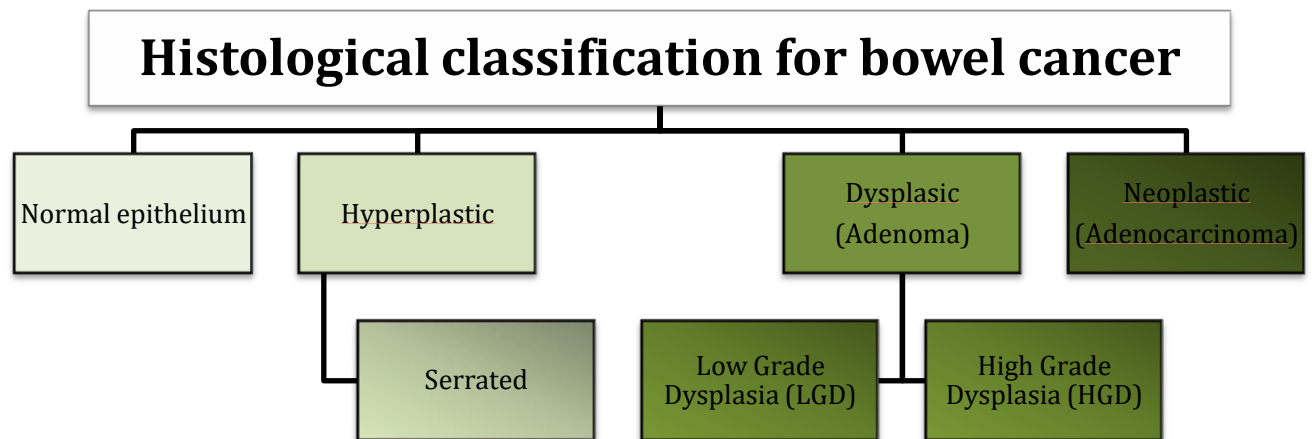
Figure 1.9: The histological classification for bowel cancer

Generally CRC is thought to arise following the accumulation of genetic and acquired epigenetic changes that transform normal glandular epithelial cells into invasive adenocarcinomas (see Figure 1.9). These transformative steps are described in the classic tumour progression model.(Fearon and Vogelstein 1990) Since this model was originally proposed, our understanding of the molecular pathogenesis has advanced considerably and led to numerous revisions of the Vogelstein & Fearon model,(Pritchard and Grady 2011) (see Figure 1.10), mostly through huge variety in genetic mutation and types of polyp. Mutations in genes responsible for DNA repair, chemical processing and the inflammatory response are all thought to have a role. For example, serrated polyps are associated with microsatellite instability (MSI) and abnormality in DNA methylation whereas adenomas arise through inactivation of APC tumour suppressor genes and display chromosome instability.(Noffsinger 2009) Although several of the molecular changes have been defined, many of the steps in the progression to cancer are still unclear and the source of much ongoing research.
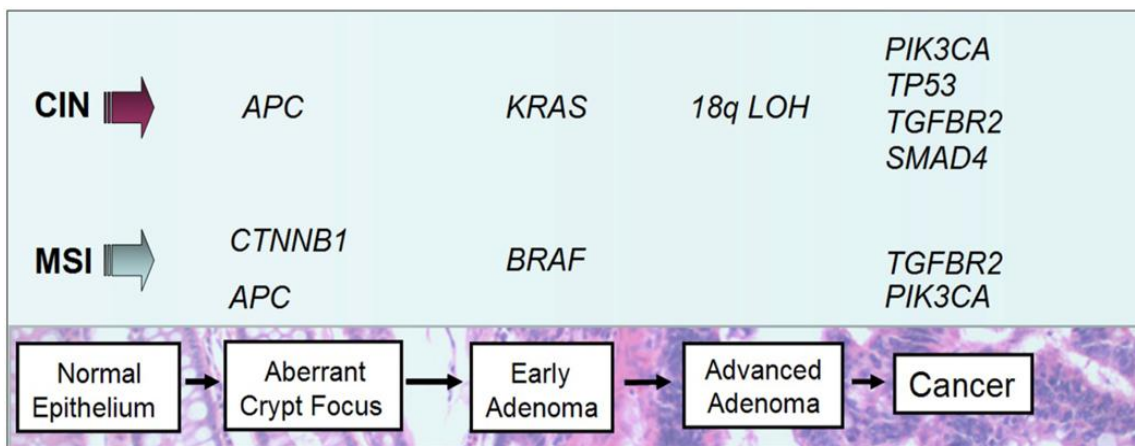
Figure 1.10: A modified adenoma-carcinoma progression model with acquired genetic abnormality at each stage of advancement(Fearon and Vogelstein 1990)

It is important to understand the molecular biology of cancer progression and its correspondence with histological change in the cells so that they can be targeted at an early stage, to try and modulate or prevent the transition of these cells from benign to cancerous.

1.6 Colorectal polyps

Defined as any lesion raised above an epithelial surface, there are many types of large intestinal polyp, which require careful assessment to ensure accurate diagnosis by the histopathologist. Polyps resected at endoscopy represent a significant workload for the histopathology team, and this has increased with the advent of the BCSP and its surveillance protocols. Significant diagnostic disagreement has been demonstrated between histopathologists for diagnosis of specific pathological entities(Wong et al. 2009) and has prompted the development of various research techniques to aid this process. Polyps can originate from both mucosal and submucosal tissues. They can be classified according to size, shape and histology.

1.6.1 Hyperplastic and other serrated polyps:

Hyperplastic polyps account for the greatest number of large bowel polyps and are commonest in the sigmoid colon and rectum. They account for about 30% of all polyps detected in the bowel cancer screening programme in the UK. They appear endoscopically as pale nodules of the colorectal mucosa and are generally small, less than 0.5cm in diameter, and sessile. They show some biochemical similarities to large bowel carcinoma but the current belief is that they represent little or no neoplastic potential. (Shepherd 2002) They do not usually display the critical abnormalities in APC and p53 gene mutations that predispose to colorectal cancer. They show a characteristic serrated phenotype, a morphological change shared with other less common polyps found especially in the right colon. The latter appear to have more neoplastic potential than standard hyperplastic polyps of the left colon and rectum and may demand enhanced surveillance strategies.

1.6.2 Adenomatous polyp:

The adenoma itself is classified as a benign polyp of the large bowel. However because the adenoma-carcinoma is thought to be the most important pathway for development of (sporadic) colorectal malignancy, identification and removal of adenomas has become the high priority during colonoscopy(Public Health England 2013). Most carcinomas arise from a pre-existing adenoma although most adenomas *will not* become malignant.   Adenomas are dysplastic by definition and display disordered, neoplastic epithelial growth patterns with failure of maturation and differentiation. Morphological features change with grade of dysplasia (worsening; mild to moderate to severe, or low to high grade) with increasingly enlarged, spheroidal and hyperchromatic nuclei with disorientation of the nuclei to the basement membrane 'polarity'.(Shepherd 2002; Clouston and Walker 2013)

 They are classified according to histological morphology: tubular (most common), villous and tubulovillous. The most important factors for predicting the development of malignancy are size, grade of dysplasia (low or high) and the villousness (the morphology of the polyp). Endoscopically adenomas are obvious above approximately 8mm in size because they have a rich vascularity

and appear red and lobulated, their exact shape dependent on the villousness and whether they are sessile or pedunculated (see Table 1.2). Adenomas can also be flat. These tend to be found in younger patients and exhibit higher grades of dysplasia and are more likely to harbour foci of invasive malignancy.(Clouston and Walker 2013) They are also less likely to be identified via the BCSP because they bleed less readily and are therefore not picked up on FOB testing. There are also various genetic disorders where patients display great number of polyps at a young age and are therefore at much higher risk of developing a colorectal cancer.

Pedunculated adenoma	**Semi-pedunculated adenoma**
Sessile adenoma with lobulated appearance	**Flat adenoma**

Table 1.2: Various presenting shapes of adenoma(Clouston and Walker 2013)

1.6.3 Mixed polyps:

These are polyps which demonstrate features of both adenomatous polyps and serrated polyps. They make up less than 1% of all adenomatous polyps and represent a polymorphous population of polyps, including collision lesions, where two forms of polyp have collided and joined.(Clouston and Walker 2013)

1.6.4 Carcinomatous Polyps:

Defined as a polyp in which cancer has invaded by direct continuity through the muscularis mucosae into the submucosa (usually on a background of low/high grade dysplasia), this can either be as a focus of invasive cells often in the centre of the polyp invading into the stalk, or the entire polyp may be composed of adenocarcinomatous tissue (having once had an adenomatous component but since destroyed by the cancer). They demonstrate architectural and cytological abnormalities within their epithelium and clear evidence of invasion with an appropriate stromal reaction (in the form of tumour-associated fibrosis or desmoplasia). As previously stated, the prevalence of adenocarcinoma in an adenomatous polyp is dependent on size, grade of dysplasia and villosity of the pre-existing adenoma. Thus, large carpeting villous adenomas have the highest likelihood of harbouring foci of invasive malignancy. In one study, 6.4% of right colonic adenomas showed malignant change, 8.0% of left sided colonic adenomas showed malignancy, whereas 23% of rectal adenomas showed malignancy.(Nusko et al. 1997)

Less than 5% of all adenomas actually demonstrate evidence of invasive malignancy and they are usually well delineated, without evidence of vascular spread and can often be effectively treated by local excision (this is polypectomy at time of colonoscopy or Trans Anal Endoscopic Mucosal Resection (TEMS procedure) (if low rectal). The endoscopist may also choose to tattoo the site of the polyp at the time so that the area is easier to find subsequently if necessary. Indeed, in BCSP, all endoscopists are mandated to tattoo the site of a polyp over 10mm in diameter because of the potential risk that that polyp may be cancerous and thus require subsequent surgery. Macroscopically carcinomatous polyps are more likely to be irregular with

bleeding from surface ulceration and firmer to palpation, and this becomes more obvious as the cancer progresses to be larger and more advanced, but these overt signs are often absent and it can be very difficult to tell an adenoma from and early polyp cancer at the time of colonoscopy.

The initial management for most polyp cancers and adenomas, therefore, is endoscopic resection and histopathological analysis of the specimen removed. Further surgical management with radical resection is fairly contentious; the evaluation of the margin of excision in polypectomy specimens is done by measuring the distance from the deepest invading tumour cells to the cauterized/excised margin off the glass slide. In well orientated intact polypectomy specimens with polyp cancer, the pathologist will also provide the level of invasion using Haggitt classification for pedunculated polyps or Kikuchi/Kudo for sessile polyps including measurements of the depth and width of the invasive component. Margin involvement, vascular invasion, poor differentiation and depth/level of invasion may indicate the need for surgical resection,(Riddell et al. 1983) which aims to prevent local recurrence and harvest lymph nodes for evidence of involvement (and therefore possibility of distant metastases) and assess the TNM status.(Haggitt et al. 1985)
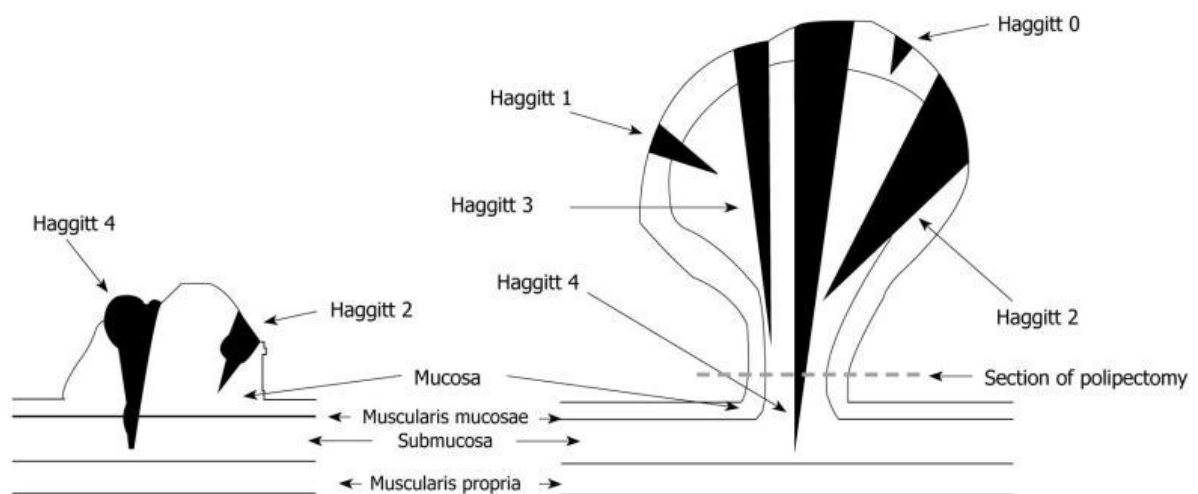
Figure 1.11: Haggitt classification of invasive cancer in a sessile and pedunculated polyp(Bujanda 2010)

Depending on the overall pathological assessment, the status of the excision margin and the level of invasion of the polyp cancer, the Multi-Disciplinary Team will make a decision whether resection of the area of the polyp is required. This may involve a major operation with a significant risk of morbidity and mortality to the patient. It is extremely important, therefore, that a correct and early diagnosis is made within an appropriate time-frame for challenging diagnostic cases, with an ultimate aim to prevent the spread of cancer, and direct subsequent management of the patient.

1.7 Epithelial Misplacement

Epithelial misplacement (EM) is a histological phenomenon which quite literally describes the misplacement of epithelium from the mucosa into the submucosal region away from its place of origin. It was first described in detail as part of a characterisation of the genetic syndrome called Peutz-Jeghers syndrome, whereby mucocutaneous pigmentation is associated with "hamartomatous" polyps. A rare autosomal dominant condition, this disease affects approximately 1 in 100,000 in the West. Patients display a mutation of the serological thymidine kinase (STK1) gene, resulting in polyps throughout the gut, nasopharynx and urinary system, but most commonly in the small intestine. Large polyps subject to high forces from gut motility were postulated to be likely to be subject to varying degrees of trauma, causing bleeding and translocation of the epithelium. The resulting EM in these cases was often dramatic and involved all layers of the bowel wall, including in the subserosa, where mucus cysts were seen.(Shepherd et al. 1987) The histological features observed demonstrated considerable mimicry of adenocarcinoma, so much that the early literature rates of small intestinal carcinoma in Peutz-Jeghers syndrome may have been falsely uplifted due to the misdiagnosis of EM.(Bartholomew et al. 1957; Shepherd et al. 1987) Misplacement of dysplastic epithelium in Peutz-Jeghers polyps has more recently been described as well, further complicating the diagnosis.(Petersen et al. 2000)

Evidence of EM has also been found in inflammatory polyps at the anorectal junction, associated with diverticulosis, and at the site of stomas and in association with mucosal prolapse, all involving large polyps under high tensile

strain. EM has also been identified in cases of serrated polyps, small hyperplastic polyps of the left colon and rectum, and in sessile polyps of the right colon, known as inverted hyperplastic polyps, perhaps now better classified as sessile serrated lesions with epithelial misplacement.(Sobin 1985; Shepherd et al. 1987; Yantiss et al. 2001)

1.7.1 Epithelial Misplacement in colonic polyps:

EM was first described by Muto and colleagues in adenomatous polyps of the colon.(Muto et al. 1973) It was noted that EM was particularly prevalent in polyps of the sigmoid colon, and they were the first to characterise the associated morphological features. Colonic adenomas often show features of misplaced epithelium in the submucosal stalk ("pseudoinvasion"), particularly in those that are large and located in the sigmoid colon.(Yantiss et al. 2002) Stalked polyps were also thought to be subject to high propulsive force in the colon, in particular in those patients with another common benign condition of diverticular disease (a totally unrelated pathology of the (usually) sigmoid colon associated with diverticula; outpouchings of the gut wall, with thickening of the muscularis propria muscle layer) which causes the internal propulsive forces to be even higher. These forces within the colon, to a greater or lesser degree, cause trauma to stalked polyps by torsion. The proposed mechanism of the pathology is initially driven by ischaemia through a twisting movement of the polyp which cuts off its blood supply and then the surface consequently ulcerates and epithelium is traumatically moved from the surface of the polyp into the central volume. It is the process of recurrent episodic ischaemia and surface damage that is thought to frequently lead to the phenomenon of EM.(Muto et al. 1973)

EM is characterised by the misplacement of surface epithelium into the central core of the polyp i.e. epithelium is seen in normal submucosa where it is not characteristically found. It is this feature that mimics invasive adenocarcinoma, in which the neoplastic epithelium migrates from its surface position because it is *invading* the submucosa. The epithelium in the submucosa of the adenoma typically appears cytologically similar to that of its surface and is usually accompanied by lamina propria tissue. Haemosiderin deposits are often

observed, in keeping with a traumatic cause for its presence, with epithelial necrosis and haemorrhage leading to that pigment accumulation. Mucus lakes are also classically described as the epithelium continues to produce mucus from goblet cells now buried under the surface and this has no way of being expelled and so it collects in large characteristic pools.

EM in adenomas may resemble adenocarcinoma infiltrating the stalk of the polyp, and it can be very difficult to tell the difference between the two when assessed by a histopathologist, and impossible for the endoscopist at time of colonoscopy. Histologically the two often have very similar features (including glandular epithelial abnormalities, the shape of glandular nests, and the presence of cystic gland and lamina propria, and haemosiderin deposits) which can complicate the diagnosis.(Shepherd and Griggs 2015) There may be evidence of acute necrosis of the surface of the adenomatous polyp, underpinning the traumatic mechanism of EM. Isolated glands without the accompanying lamina propria, tumour budding, evidence of vascular invasion and poor cell differentiation are all features that favour the presence of invasive adenocarcinoma.

1.8 Clinical Diagnostic Need

Increasing numbers of cases of EM have been seen in recent years which are attributed to the advent of the BCSP in recent years. It is thought that because sigmoid polyps are predisposed to bleeding, more patients are preferentially picked up via the faecal occult blood testing (at home testing) and therefore more polyps are being found at colonoscopy and removed and sent for histological analysis. It is likely that these polyps will continue to be seen with increasing frequency (Clouston and Walker 2013) and cause more of a problem for the diagnosis of polyp cancers for clinicians. There are no endoscopic, histopathological or clinical methods to reliably differentiate cancer from EM, but clearly it is extremely important to correctly identify the cancer cases from those of EM because of the clinical implications, difference in patient management (polypectomy versus colectomy)(Hurley et al. 2013), and patient reassurance. Pathologists only have routine histopathological processing of tissue (with stains) to make the diagnosis. This traditionally uses special stains to identify

cells or areas of interest in the tissue based on contrasting colours within a sample and are, in the most part, reliable. Immunohistochemistry for commonly used markers including E-Cadherin, Collagen IV, MMP-1 and p53 have been promoted(Mueller et al. 1997; Hansen et al. 1999; Yantiss et al. 2001; Yantiss et al. 2002) to aid diagnosis. The morphology of the tissue cells and tissue architecture can be visualised and from this a diagnosis is made, although these do not always confer benefit for the difficult intermediate lesions.

Analysis of tissue using traditional staining techniques is technically challenging, time-consuming and user-dependent, as with most acquired skills it is generally reliable in cases with some obvious pathology, however can be very limited in difficult cases where the pathology of one disease state approaches the other.(Clouston and Walker 2013)

Recently described as the "diagnostic conundrum of the century",(Shepherd and Griggs 2015) it is sometimes so difficult to differentiate between EM and a focus of carcinoma in an adenomatous polyp that the local hospital GI Histopathologist is unable to settle on a diagnosis. Diagnostic agreement at both regional and national level between expert pathologists has not always been unanimous(Turner et al. 2013) and in 2009 the BCSP established a National Expert Board (EB) to assess the problem and to address difficult diagnostic cases. Funded by the BCSP in England, over 200 individual cases have so far been through the EB. The board compromises three GI Registry Consultant Histopathologists who each review all the cases blinded. The original slides are sent to each Consultant in turn and they each provide a diagnosis and send on to the next. An overall diagnosis is made by majority consensus and a letter written back to the original pathologist. Even so, there are cases where the EB has been unable to make a definitive diagnosis because one or more expert pathologist has provided an 'equivocal' diagnosis. Although absolutely vital, this process is obviously time-consuming, expensive and as a consequence may delay diagnosis for the patient and onward management decisions.

CHAPTER 2

VIBRATIONAL SPECTROSCOPY

2.1 Biophotonics

Biophotonics is the combination of biology and photonics; concerning generation, emission, transmission, modulation, detection, amplification, and sensing of light.(Bellisola and Sorio 2012) Biophotonic technologies can objectively measures the changes associated with cancer progression, such as change in genes, proteins, metabolites, morphologic, and other detectable traits.

Spectroscopy is specifically the measurement of the interaction between radiated energy and matter. Vibrational spectroscopy, incorporating both infrared (IR) and Raman spectroscopy, is the analysis of the spectra produced by the interaction of light with the vibration states of molecular bonds. They provide information about the structure of a molecule.

2.2 What is light?

Faraday proposed in 1847 that light was a high-frequency electromagnetic vibration, which could propagate even in the absence of a medium such as the ether. James Clark Maxwell found that self-propagating electromagnetic waves would travel through space at a constant speed, which happened to be equal to the previously measured speed of light, and in 1862 proposed that light was a form of electromagnetic radiation (EMR) in wave formation (See Figure 2.1).
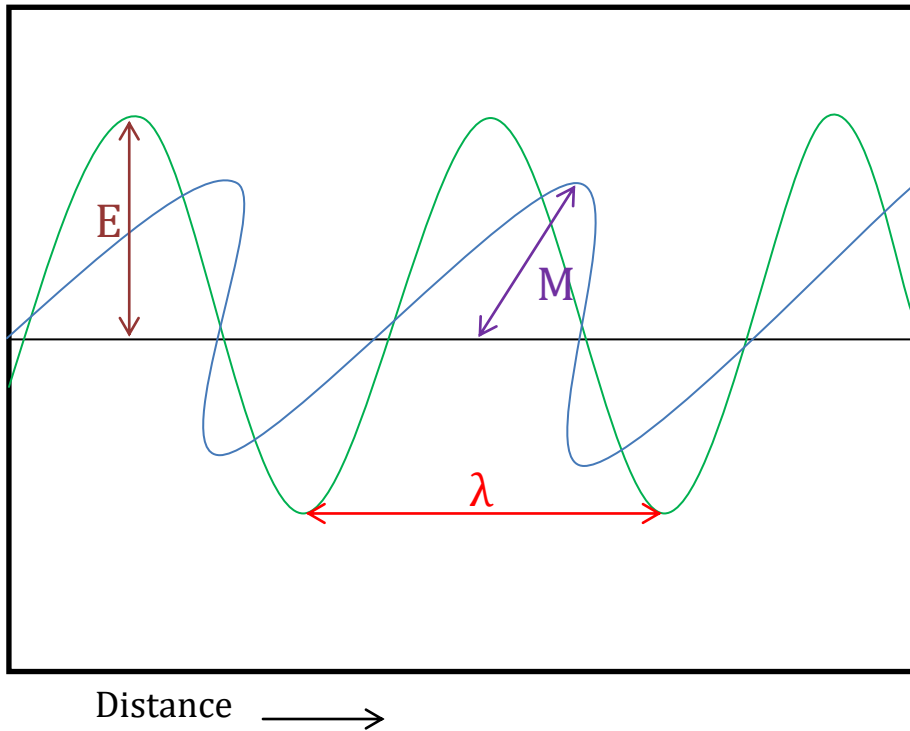
Figure 2.1: Description of a light as a wave

The frequency of the wave determines its colour: $4×10^{14}$ Hz is red light, $8×10^{14}$ Hz is violet light (wavelengths in the range of 700 nm ($700×10^{-9}$ m) to 400 nm respectively), and in between these are all the other colours of the rainbow in visible light– between the infrared, with longer wavelengths and the ultraviolet, with shorter wavelengths. Generally, EMR is classified by wavelength into radio, microwave, infrared, the 'visible light' region, ultraviolet, X-rays and gamma rays, see Figure 2.2.

Increasing Frequency

γ rays | X rays | UV | IR | Microwave | AM | FM | Long radio waves

Radio waves

Increasing Wavelength

400nm

700nm

Visible spectrum
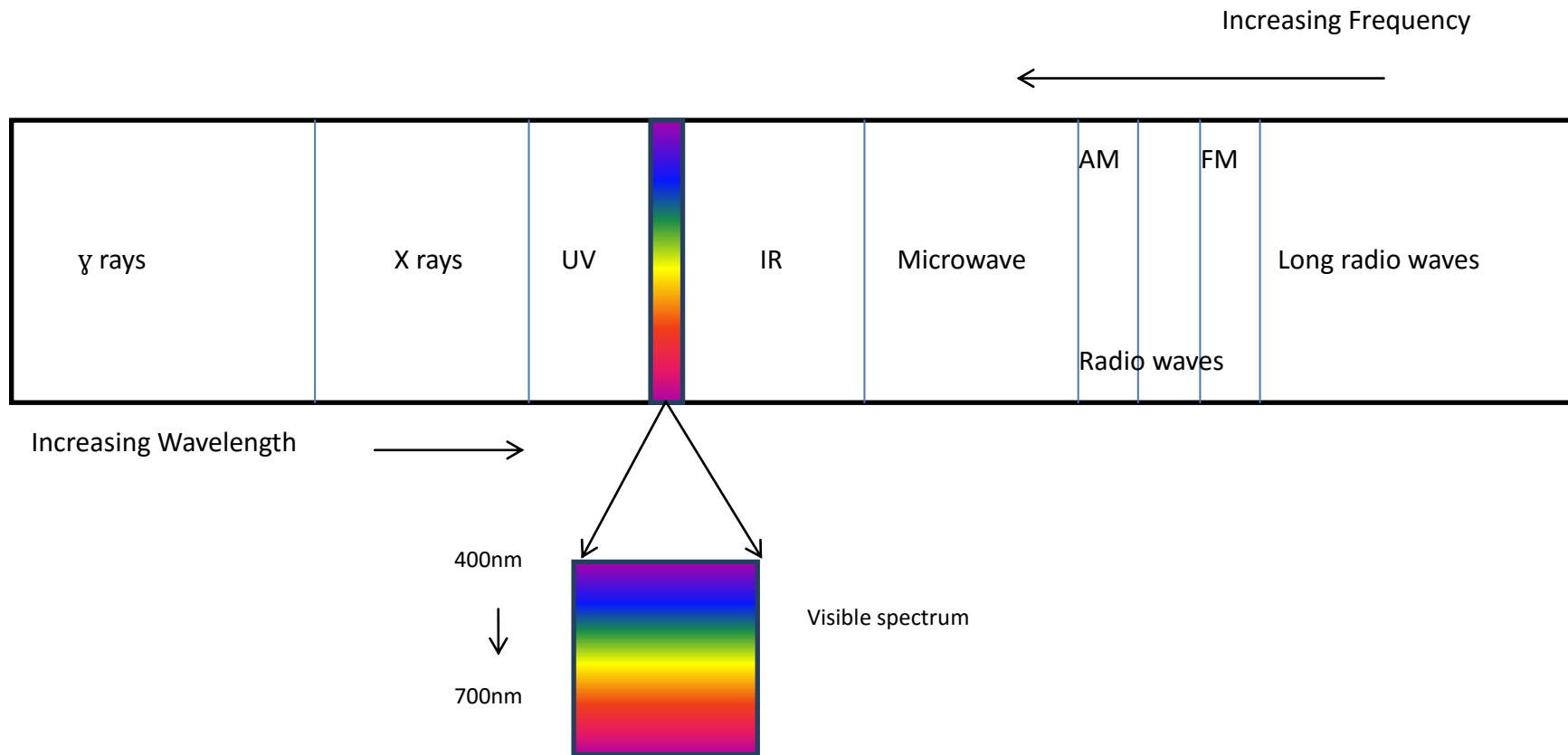
Figure 2.2: A schematic representation of the whole electromagnetic spectrum

The speed of light in a vacuum is defined to be exactly 299,792,458 ms^{-1}, approximately 186,282 miles per second (the meter is now defined in terms of the speed of light, hence approximate value in miles per second). All forms of electromagnetic radiation move at exactly this same speed in vacuum giving them wavelengths inversely proportional to their frequencies. In 1900 Max Planck suggested that although light was a wave, these waves could gain or lose energy only in finite amounts related to their frequency, known as quanta. In 1923 Arthur Holly Compton showed that the wavelength shift seen when low intensity X-rays scattered from electrons (Compton scattering) could be explained by a particle-theory of X-rays, but not a wave theory. Thus eventually light was described in terms of *both* a particle and a wave. Radio waves and X-rays involved in Compton scattering tend to behave more like a classical wave at lower frequencies, but more like a classical particle at higher frequencies, but never completely loses all qualities of one or the other.

EMR in the visible light region consists of photons capable of causing electronic excitation within molecules, which lead to changes in the molecular bonding or chemistry. At the lower end of the visible light spectrum, EMR becomes invisible to humans (infrared) because its photons no longer have enough individual energy to cause a lasting change in conformation in the visual molecule retinal in the human retina.

2.3 Biological Diagnostics

There are many radiological imaging techniques that aid clinician in making a diagnosis of an abnormality in the gut (computer tomography, barium enema, capsular endoscopy, magnetic resonance scan) however these do not allow concurrent tissue diagnosis and patients have to be brought back at another time to take biopsies of suspicious lesions. Colonoscopy, as described in Chapter 1, remains a sensitive method of imaging the entire colon, with the benefit of tissue diagnosis and treatment with tissue resection in some instances, and is considered to be the Gold standard in 'screening' or detection tool. Colonoscopy enables the clinician to visualise pathology inside the gut and

to acquire tissue biopsies for further analysis at the same time. Further techniques such as tattooing are also available so that the area can be identified at a later stage, which make colonoscopy as a technique a very efficient process with satisfaction for the patient and clinician.

The current gold standard for diagnosis of tissue abnormalities, specifically cancer and early change of cancer, remains with conventional histopathology. Specimens taken at colonoscopy; either whole polyps or in piecemeal are fixed in formalin and sent to histopathology lab for processing and assessment. This involves further tissue manipulation embedding it into paraffin, sectioning and then staining the sections of tissue. They can then be visualised by the pathologist using a microscope. This is a time-consuming process, and preparation is very dependent on the skill of the practitioner. The main problem with individuals reporting cases is the subjective nature of interpretation. There is often a degree of inter-observer disagreement(Lieberman et al. 2000; Turner et al. 2013) and this has proved to be an ongoing problem for clinicians to make an absolute, or definite diagnosis especially in some cases where several pathologies are very similar. Colonoscopy itself is considered to be a fairly low risk procedure- it is a day-case procedure in the UK and patients arrive at the hospital on the day of their procedure and go home the same day, all being well, with only a short period of monitoring necessary after (and this is often dependent on the sedation drugs used during the procedure). The main risks are either immediate risk to the patient of bleeding during the procedure and perforation (following polypectomy or biopsy during the procedure or from the scope passing around and tearing the bowel) although sometimes this is not diagnosed until after the procedure. Also very small polyps can go unnoticed by the clinician and these therefore do not get removed at the time but have the potential to develop into a cancer. Small flat (particularly right-sided) adenomas can be easily missed.(Horiuchi and Tanaka 2014) Delay in organisation of colonoscopy by department can cause patient distress, and incur high financial and resource costs to the hospital.

There is therefore a real need for complimentary diagnostic technology during or adjunct to the colonoscopy procedure to help simplify and speed up diagnosis. In recent years several biophotonic diagnostic methods that are complementary to conventional histopathology have been developed to reduce

the subjectivity and invasiveness of diagnosis in the colon. A range of enhanced optical techniques have been developed to aid clinicians and pathologists in diagnosis and efficiency of routine clinical practise.

These mostly aim to assist/ improve in one of the following four clinical areas;

- greater polyp detection
- optical biopsy
- identification of flat dysplasia
- risk stratification through field carcinogenesis detection.(Roy et al. 2011)

2.4 Overview of Wide-field Imaging Techniques

2.4.1 High definition white light endoscopy

High definition white light colonoscopy is gradually replacing the traditional white light because the images are clearer and crisper and generally better quality. However there is no substantial evidence that this actually cause any improvement in adenoma detection rate- a large meta-analysis in 2011 demonstrated only a 3.5% improvement in detection rate in the high definition group.(Subramanian et al. 2011) The equipment is more expensive and it has been difficult to justify rolling out the new high definition sets for little perceived gain, however often units are replacing their old scopes and stacks with the new technology as they become worn out and fail.

2.4.2 Chromo-endoscopy

Indigo carmine spraying of colonic mucosa through the instrument channel of the colonoscope increases the visual contrast between normal and abnormal mucosa and therefore makes it easier to pick up lesions than colonoscopy alone. It is a simple, safe technique that uses current colonoscopes. A randomised control trial in 2011 demonstrated an increased detection rate of adenomas in the chromo-endoscopy group as opposed to the white light colonoscopy, however it was noted to significantly increase the overall length of the procedure which was again, operator-dependent.(Pohl et al. 2011) It was noted to be helpful in visualising the boundary of a lesion and defining very small lesions better. However spraying of the entire colon was so time-

consuming that it was thought to be impractical for everyday use. It was a technique particularly popular in the Far East as these clinicians have a lot of experience with the technique in the upper gastrointestinal tract.(Song et al. 2007) A classification of polyps using a magnified view of the colonic crypts (up to x 100) and indigo carmine dye was developed called the Kudo classification. It divides polyp pit pattern into 6 categories which correlate to histological findings and help the operator to determine if a lesion is likely to be malignant.(Kudo et al. 1994; Liu et al. 2003)

2.4.3 Virtual Chromo-endoscopy

Narrow band imaging (NBI), I-Scan and Flexible Spectral Imaging Colour Enhancement (FICE) use optical/ electronic methods to gain more detailed mucosal surface information. NBI is a diagnostic imaging technique where the mucosal detail is enhanced by the application of blue and green light that is activated by a switch in the endoscope. These wavelength bands correlate to the peak light absorption by haemoglobin and therefore vessels containing blood appear darker. NBI allows the morphological and structural character of lesions to be examined in greater depth, including the pit pattern, without the need for dye. A comparison of techniques is seen in Figure 2.3, demonstrating an increase in degree of polyp surface detail obtained using carmine dye and NBI compared with simple white light colonoscopy. NBI has been shown to demonstrate a diagnostic accuracy of over 95% when comparing malignant and non-malignant lesions.(Sano et al. 2009) The pickup rate for abnormalities equals chromo-endoscopy, with the benefit of better visualisation of the vascular pattern, without the need for time-consuming spraying and no need for the dye itself. Based on colour, architecture, and pattern, NBI enables a more accurate diagnosis at time of colonoscopy without the use of dye.(Tanaka and Sano 2011) The NBI system is again based on the subjective judgement of the endoscopist, so recent research has been directed at developing an objective, in vivo real time diagnostic tool to distinguish adenoma from carcinoma.(Inomata et al. 2013)
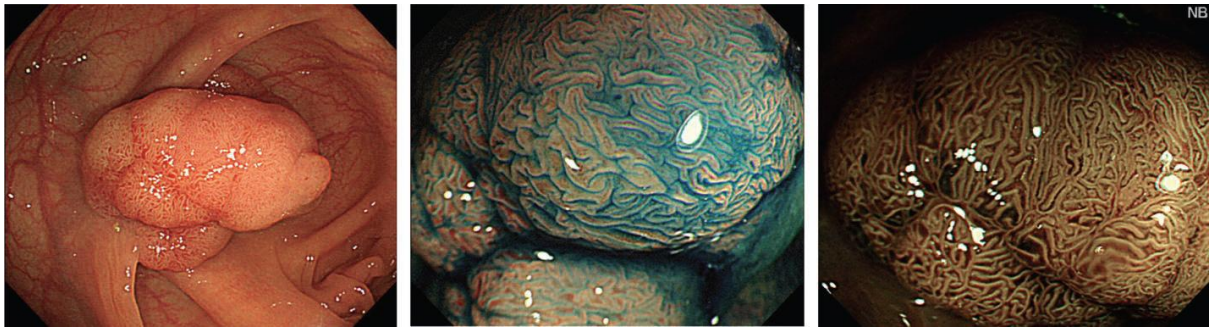
Figure 2.3: Adenoma detection with (Left:) Simple white light coloniscopy, (Centre:) Carmine dye enhanced, and (Right:) narrow band imaging (Tanaka and Sano 2011)

2.4.4 Fluorescence spectroscopy

Fluorescence spectroscopy as the name suggests, uses tissue fluorescence to differentiate tissue types. Fluorescence occurs when light (of specific wavelength for fluorophore) is directed onto a tissue; the light is absorbed and causes autofluorescence from endogenous molecules within such as collagen, elastin and various enzymes (the fluorophore). The ratio between incident light and fluorescence intensity can be used to discriminate various tissues depending on the content of endogenous fluorophores present.(Taylor et al. 2007) It is also possible to use exogenous molecules, supplied either topically or systemically, which increase the production of fluorophores within abnormal tissue and enhance the contrast with normal tissue. It is particularly interesting and useful in areas of flat dysplasia, alone or within inflammatory pathology but there have been significant false positives with this. The majority of in vivo work by various groups has been done with laser-induced fluorescence spectroscopy (LIFS) either utilising natural fluorophores or exogenous drugs that target specific tissue. This is less favoured because of potential patient-drug interactions. LIFS to detect autofluorescence has been used successfully in brain, bladder and oral cavity but sensitivity has been unacceptably low with a significant number of false positives, particularly in the bladder.(Andersson-Engels et al. 1990; Svanberg et al. 1994; Vo-Dinh et al. 1995)

The difficulty over the last few years has been actually proving that these techniques significantly improve the detection of adenomas or diagnosis of

cancer. A randomised controlled study in 2011 combined high resolution endoscopy with NBI and AFI in a single endoscope called trimodal imaging (ETMI) and compared with standard video endoscopy. They were overall unable to improve adenoma detection rate although found ETMI was useful in discriminating adenomas from normal tissue.(Kuiper et al. 2011) Similar work has been replicated a number of times and only minimal improvement has been demonstrated. Consequently in most hospitals these techniques are only employed by clinicians specially trained or dealing with particularly high risk surveillance groups. In the last few years, a team at Washington State University has been developing the use of real time intra-operative near-infrared fluorescence goggles. The patient is injected with indocyanine green which is taken up by the tumour cells pre-operatively and the surgeon wears the goggles during the procedure to aid the removal of the whole of the tumour (a benefit when the edges are not well demarcated). The research was first trialled in rat models but is now in human trials phase and hit the newspaper headlines earlier on in the year.(Liu et al. 2013; BBC Health 2014) This technology could potentially be applied for use during colonoscopy for real-time tumour diagnosis and margin assessment during polypectomy, however it does not currently confer the subtlety needed to differentiate between similar pathologies in the gut.

2.5 Point Measurement Techniques

2.5.1 Confocal Fluorescence microscopy

Confocal microscopy is a non-invasive method of optical imaging in two different ways; reflectance, based on the difference in refraction of cellular tissue structures and fluorescence, based on natural fluorophores in the tissue, both *in vivo* and *in vitro*.(Ragazzi et al. 2014) An image is formed by detection of high-resolution fluorescence generated from a scanning laser beam, providing optical slices through the sample and creating a 3 dimensional image when a whole stack of these are obtained. Keisslich *et al.* performed the first colonic studies and used fluorescent agents to obtain images with a sensitivity of 97% and accuracy of 99% when compared with histological assessment.(Kiesslich et al. 2004) However the need for an exogenous fluorophore made this unpopular

and experimental difficulties were found with peristalsis of the gut and debris interfering with the signal.

2.5.2 Light scattering spectroscopy

Light scattering spectroscopy is based on the elastic scattering of light (without a change in energy level). Tissue pathologies are detected and diagnosed using spectral measurements of the elastically scattered light, in a manner that is sensitive to both scattering and absorption properties of the tissue, over a wide range of wavelengths.(Mourant et al. 1996) Particles scatter light best when the wavenumber approaches their own size; neoplastic tissues have increasingly large amounts of nuclei and mitochondria when compared with normal so the elastic scattering is of a longer wavelength.   A study in 2006 by Dhar *et al.* used the technique with an *in vivo* probe and found the technique to be promising in the differentiation of normal, inflammatory and neoplastic tissue. The aim was to develop a real-time diagnostic implement that could potentially reduce the unnecessary removal of certain polyps and therefore minimise the risk to the patient, the results however, were not found to be particularly sensitive.(Dhar et al. 2006)

2.5.3 Optical Coherence Tomography

This is a non-invasive imaging technique that relies on light scattered by the sample and subsequent interference, and provides a cross sectional image of the tissue(Fujimoto et al. 1995) an example of which can be seen in Figure 2.4. Light is reflected from surface tissue microstructures and detected to a spatial resolution of 4-20 μm.(Sivak et al. 2000) The resolution is many more times greater than ultrasonography, an analogous technique, and so the images created are much more detailed, and based on cellular architecture rather than tissue structure. Visual inspection of OCT images can be used to assess tissue morphology, and to differentiate between normal and abnormal tissue types.(Gossage et al. 2003)
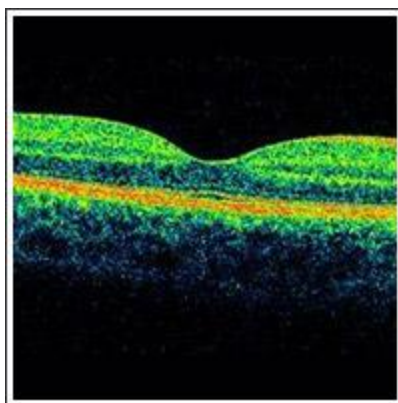
Figure 2.4: An example of Optical Coherence Tomography used to examine the retina of the eye in detail(Eye Site Texas 2015)

Studies have combined the use of Raman and OCT as a useful tool for tissue diagnosis combining the molecular and topographical information obtained. Ashok et al. combined the two in comparison of normal and cancerous colonic tissue in vitro and found that the combined approach improved sensitivity of detection from 77 and 74% respectively to 94%.(Ashok et al. 2013) However the technique is limited by depth of investigation and so its use for real-time diagnosis is limited because of this in colonoscopy, other than perhaps to assess the depth of tumour invasion through the muscularis mucosa, which may help staging for cancer. It is also a very time-consuming technique because of the length needed for image acquisition and operator dependent.

2.6 Biospectroscopy

Infrared (IR) and Raman spectroscopy are techniques used to interrogate biological tissues in order to understand the changes in the molecular constituents that are associated with cancer progression from normal to malignant. These spectroscopic techniques have the ability to provide detailed biochemical information since the spectrum obtained from cellular tissue measurements are intrinsic molecular fingerprint of the sample that reveals information about the DNA, carbohydrate, protein and lipid content.(Chan et al. 2006) Each spectrum is made up of varying contributions from many molecular bond vibrations and is therefore a superposition of all constituents in the

measured sample which therefore contains potentially useful biodiagnostic information.(Naumann 2008)

2.6.1 Raman Spectroscopy

Raman spectroscopy is the measurement of inelastically scattered light. It requires monochromatic irradiation and detects a scattering pattern from the excitation of molecules.(Lyng et al. 2015) White light incident upon a specimen may be altered in wavelength and this can be measured and represented as a Raman spectrum. This can provide a unique biochemical 'fingerprint' of the sample.(Kendall et al. 2009) The largest and most recent Raman study in the colon was by Wood et al. in 2014; a modified confocal Raman probe was used to differentiate cancer, inflamed and normal tissue with a sensitivity of over 75%.(Wood et al. 2014) There is huge precedence of the use of Raman probes *in vivo* in the oesophagus, where a variety of different probes with different design in different tissues have been used.(Almond 2012; Kallaway et al. 2013) Future work needs to be performed to use the Raman probe *in vivo* in the colon at time of colonoscopy to make real-time diagnosis, particularly of microscopic colitis which could lead immediate treatment for the patient.

2.6.2 Infrared Spectroscopy

Infrared spectroscopy is the absorption of photons detected when molecules in a substrate are irradiated with polychromatic mid-infrared light. In this study we propose to use infrared spectroscopic analysis of colon cells to aid the difficult diagnosis of epithelial misplacement and attempt to streamline the diagnosis of cancer.

CHAPTER 3

INFRARED SPECTROSCOPY

Infrared (IR) radiation spans an interval of the electromagnetic spectrum between the red end of the visible region (wavelength ~780 nm, frequency ~3.8×10^{14} Hz) and the beginning of the microwaves region (1 mm, 3×10^{11} Hz). The infrared band can be split into:

- Near – IR (from ~780 nm to 2.5 μm)
- Mid – IR (from 2.5 to 25 μm)
- Far – IR (from 25 to 1000 μm)

Within the infrared band, the light waves vibrate at varying frequencies, which is inversely proportional to (sinusoidal) wavelength, λ, when c is constant speed (phase speed of the wave):

$$\lambda = c / v$$

And the most commonly used spectroscopic unit in the IR region is the reciprocal of the wavelength, 1/λ, called the wavenumber.

Molecular vibrations can be excited by the absorption of light.(Herzberg 1950) When infrared incident light is shone on a tissue, the molecules absorb photons which results in change in the vibrational mode. A molecular vibration occurs when atoms in a molecule are in motion, while the molecule as a whole is also in motion (translation and rotation). The frequency of the atomic motion is the vibrational frequency. A molecular vibration is excited when the molecule absorbs a quantum of energy, E, corresponding to vibrational frequency, v:

$$E = hv \qquad \textit{(h is Planck's constant)}$$

When a sample is placed in the path of an IR beam between the source and the detector, the molecule will absorb the frequency of IR that coincides with the frequency of vibration that the molecule enters in a resonant vibrational state. The change in energy through absorption alters the *dipole moment* of the molecule, thus the charge distribution of the molecule is changed.

3.1 Vibrational Modes

Atoms in a molecule are constantly moving while the whole molecule experiences continuous translational and rotational movement. A molecule with two atoms has just a single motion, but molecules with any number of atoms have more than one type of vibration, or mode.

The vibration of molecules detected by IR spectroscopy is determined by the number and type of bond movements. IR spectroscopy detects changes in the permanent dipole (not molecules with only two atoms, unless they are asymmetric e.g. CN, because they only have one bond and therefore one single vibration). Polyatomic molecules have several normal modes of vibration which include: asymmetric, symmetric, twisting, scissoring, rocking and wagging, as seen below in Figure 3.1. The number of vibrational modes for a non-linear molecule is described as 3N-6, where 3N is the 3-dimensional movement around N atoms (or ions), and -6, rotational motions around 3 axes.

+ Above the plane   − Below the plane

Scissoring   **Rocking**   **Wagging**   **Twisting**

In-plane Bending   **Out-of-plane bending**
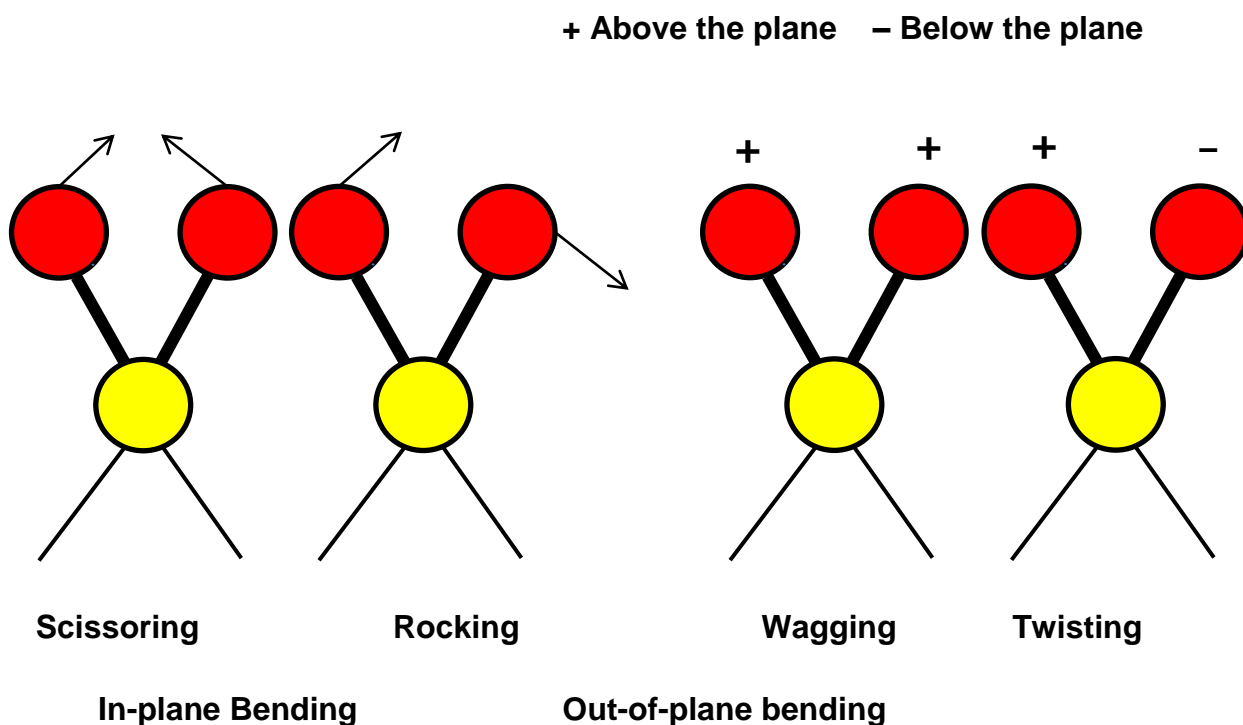
Figure 3.1: Variations of molecular bond movements

A water molecule, H-O-H, for example, exhibits asymmetric and symmetric stretching and bending vibrations at ~3500 cm^{-1}, ~1650 cm^{-1} and 600 cm^{-1}. Carbon dioxide, 0-C-0, is detected at ~2560 cm^{-1} (asymmetric) and ~500 cm^{-1} (scissoring bending).

IR spectroscopy therefore essentially measures the loss of radiation transmitted through a sample. The intensity of absorption bands can be plotted against interval of wavenumbers, which corresponds to changes of vibrational energy levels from the ground level to the first energy level in molecules ($E_0 \rightarrow E_1$)(Bellisola and Sorio 2012). The intensity in peaks reflects the amount of different molecular bands absorbing specific regions of the IR spectrum. Frequencies at which absorption at specific vibrations occur are fairly constant and can therefore be reproducibly detected within a sample, see Figure 3.2.
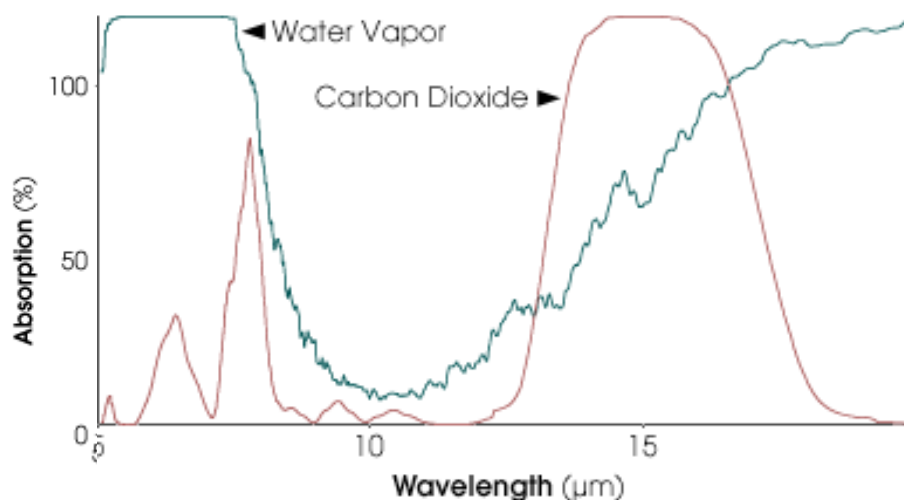
Figure 3.2: An example of infrared absorption peaks and the molecular bond attributed to water and carbon dioxide(The Earth Observatory 2014)

3.2 Clinical Need for Infrared Spectroscopy

Infrared (IR) spectroscopy offers the advantage of being able to perform the technique on samples routinely processed by the histopathology labs. Its use is more limited in fresh samples because of the presence of water molecules in

Page | 52

the tissue which are identified very readily in the IR spectra to the cost of other information. This limits the use of IR for real-time, however it is the ideal investigative tool for analysis of formalin-fixed, paraffin embedded specimens. This also confers the added benefit of multiple measurements from the same sample because there is no concern that the fixed sample will dry out, as is the case when analysing fresh samples. The water signal is also reduced in the processing of fixing so the noise is reduced. Until recently however, the paraffin signal was often very marked in results and made computer analysis of the spectra complicated. Cluster analysis is for example, very influenced by the presence of paraffin signal and can affect interpretation of the results.(Hafiane et al. 2008) However recent advances in computer programming have enabled the paraffin signal to be counteracted at the pre-processing stage; numerical techniques applied directly to the spectral map, rather than chemical deparaffinisation (thus simplifying the experimental protocol)(Nallala et al. 2012) which has made the analysis of spectra produced with this technique easier and more streamlined. It also enables samples taken during routine clinical assessment such as polyps at colonoscopy which are then sent for routine processing, to be analysed by IR spectroscopy. This creates a huge potential wealth of samples that can be analysed from historical record and with little need for further processing or need to collect more samples. There is no evidence to suggest that age of paraffin-embedded specimen affects results, so the potential role for identifying historical samples from tissue BioBank and Pathology departments is huge. IR spectroscopy is a technique that therefore lends itself to in-depth analysis of historical samples without further need for processing, and as such the technique is highly applicable for use with routine processing methodology in place within the NHS. For these reasons it was decided that analysis IR spectroscopy would be the optimum technique for analysis of historical EM samples.

3.3 A Historical Overview of infrared Spectroscopy

Infrared spectroscopy (IR) provides, in a non-destructive and label-free manner, a biochemical fingerprint of cells and tissues.(Martin et al. 2010) A spectral map is created, where each pixel of the image provides a complete spectrum across the wavenumbers(Lasch et al. 2004). This information obtained from a sample

offers insight into the presence of particular molecules and their structure and metabolism. This is of particular relevance and interest when monitoring molecular change in early disease such as the change from high grade dysplasia to invasive carcinoma. The technique, coupled with microscopy has been recognised since the mid- 20th century. However it wasn't until the late 1970s that commercially driven IR spectrometers were introduced. It gained popularity through its use of trace samples which did not get destroyed during measurements and required minimal sample preparation; results were frequent, precise and reproducible.(Bellisola and Sorio 2012) Combined with a Fourier Transform technique (a mathematical process which allows simultaneous collection of the full spectral range and converts raw data into the actual spectrum) and appropriate statistical data analysis methods, IR spectroscopy has been used over the decades as a helpful diagnostic tool in various organs including breast,(Rehman et al. 2010) oesophagus,(Quaroni and Casson 2009) colon,(Rigas et al. 1990) liver,(Le Naour et al. 2009) and lymph system.(Bird et al. 2008)

3.4 Applications of Infrared Spectroscopy

Rigas et al. demonstrated a difference in FTIR signal between normal and cancerous fresh tissue from the same surgically resected colon in 11 patients.(Rigas et al. 1990) Each tissue section was compared with stained H&E slide and examined by two blinded pathologists. Data was presented as a description of the change in the bands between normal and cancerous tissue, but there was evidence of significant inter-patient variability with such small sample numbers. They noted a change in the intensity of bands at certain wavenumbers, particularly those corresponding to phosphate and nucleic acid bands, and particularly noted a shift in the peak maximum of the bands and their centre of gravity, but no attempt was made at complex analysis further than band observations. Argov et al. were the first group to publish a direct comparison of normal and cancer spectra with those of polypoid lesions in three patients.(Argov et al. 2002) They postulated that the FTIR signal should demonstrate the changing biochemistry of early neoplastic change in adenomas. The total number of patients was included was 24, so more inclusive of population variation than the Rigas paper. They used 128 scans

and generated a total of 300 spectra for analysis. Spectral analysis was also more complex and was performed using an artificial neuronal network (ANN) to generate a prediction model and compare morphological differences between specimens. Peak analysis was performed on only 3 patients; RNA/ DNA ratio was found to increase with malignancy but was a poor discriminator of normal and adenomatous tissue, however by contrast levels of carbohydrate were higher in normal than for malignant tissue. The ANN classified 24 tissue types based on 5 features resulting in a sensitivity of 81-89% across the pathology groups, however specificity was lower.

Li et al. performed in vivo studies on colon in 2005 with the potential of assessing resection margins intra-operatively or real time pathological diagnosis.(Li et al. 2005) A small number of cases were included, total from five patients were included although one of these was rectal. They used fibre-optic attenuated total reflectance (ATR) FTIR spectroscopy in vivo and in situ intra-operatively to measure spectra on the surface of diseased tissue with a reading from normal tissue close by. The rectal tissue was actually measured pre-operatively with the patient awake, taking a minute in total to acquire 32 scans. Analysis was descriptive of the change in bands between normal and cancerous tissue, particularly of carbonyl, methyl and methylene (reduced in cancer presumably because of higher energy requirements in the cancerous tissue using local triglycerides which contain these specific groups) and all spectra correlated with histopathological result for the each patient. No attempt at more complex analysis was made however, and these results although promising, are limited due to lack of case numbers and therefore total number of spectra, variation in population and assumptions made about peak assignments without quantifying the data further.

Techniques for demonstrating the success of IR spectroscopy gathered particular momentum as a diagnostic tool in the following few years.(Krafft et al. 2009) Lasch et al. used FTIR spectroscopy to map one case of rectal carcinoma using fresh frozen tissue and compared a number of unsupervised analysis techniques.(Lasch et al. 2004) They produced a variety of false colour maps of a section of cancerous tissue and compared the images for a number of different computer techniques, which is a much more robust way of analysing the spectra than simply peak assignments. They obtained very clear images

from the three techniques and showed that this new technology could produce images with much more detailed spatial information present comparable to traditional pathology stains from which diagnosis could be achieved. Various statistical analyses began to be applied to tissue maps to increase the contrast between tissue structures in normal and cancerous tissues.(Kendall et al. 2009) Once the maps had been processed with multivariate statistical approaches, biomolecular information could be compared between normal and abnormal and the presence of tumour or inflammation in the tissue identified.(Wood et al. 2011) As such, two years after their initial results series Lasch *et al.* published similar work with supervised analysis techniques to identify great detail within the tissue sample, including neoplastic epithelium, necrotic areas of tumour and mucin.(Lasch et al. 2006)

Krafft *et al*. also produced FTIR microscopic images and compared them with those acquired from Raman spectroscopic imaging.(Krafft et al. 2008) They used a single colon specimen and snap froze the acquired tissue to analyse via the two methods. Three distinct tissue regions were compared using the two techniques and unsupervised k-means cluster analysis applied to segment the data sets, although the cluster assignments were not then evaluated for accuracy by a pathologist for Gold Standard of assessment. They assigned the clusters to various histological structures and were able to favourably compare Raman and IR for specific areas of tissue. Peak assignments of the clusters were analysed for spectral differences and described for the proposed histological areas.

Conti *et al.* compared healthy and cancerous colon tissue on formalin-fixed, paraffin-embedded using near infrared FTIR microspectroscopy.(Conti et al. 2008) They collected a bigger dataset of 1269 spectra of unknown number of patients at a resolution of 50 x 50μm in reflection mode. Band assignment was carried out in accordance with data from the literature with a change in intensity and shift of some significant peaks. Multi-variant statistical analysis was performed before creation of topographical custom colour maps using the spectra loaded onto a chemical map. They were confident that these colour maps represented the corresponding pathology on H&E slide. Ultimately a number of 'spectral patterns' were suggested to be able to distinguish normal

from cancerous tissue, however there was no statistical evaluation of these suggestions.

A large study was performed by Zweilly et al. comparing the spectra of 240 patients across 5 groups which included normal, cancer and adenoma with varying grades of dysplasia.(Zwielly et al. 2010) They also found differences in peaks at various wavenumbers which could be assigned to specific molecules and postulated that lipid/protein ratio increases with increasing severity of disease, and phosphate levels were lower reflecting the high metabolic turnover in malignant tissue. Cohenford *et al.* also performed FTIR spectroscopy on fresh frozen tissue from specially bred mouse colons and cell lines.(Cohenford et al. 2012) Classic spectral analysis of band peak intensity ratios, areas, and shifts in spectral peaks were compared for over 700 sites from 10 mice, as well as subtle differences between the spectra detected with multivariate statistical analysis. In-keeping with previous studies they found a statistically significant difference between the mean and SDs of amide II/I ratios in normal colon cells and in colon cells with high-grade dysplasia, suggestive of a variable amount of DNA, assigned to the high turnover associated with malignancy. Statistical analysis was very successful with a sensitivity of 91.4-100% and specificity of 94.1-100%. Misclassified spectra were low. Whether or not these results would be replicable in human tissue however remains to be seen.

Later experiments focused on the use of multivariate spectral analysis in various forms to analyse results and display subtle differences between pathological groups. Wood *et al.* performed FTIR spectroscopy and mapping across a range of colonic pathologies for the first time, including results from areas of inflammation in diseases such as ulcerative colitis.(Wood et al. 2011) This was also the largest study with regards to number of spectra to date (6944 from 10 samples). A 2-dimensional plot of the groups was constructed to map findings over the white light image and PCA-fed LDA and cross validation methods performed to assess the accuracy of the classification model. Prediction accuracy varied from a sensitivity of 91.6% to 100% and a specificity of 97.7 to 99.9%. Ordinary least squares relative signal contributions to tissue spectra were demonstrated following analysis of first derivative spectra to explain the biochemical basis of the accuracy of the model.

Kallenbach-Thieltges *et al.* used FTIR in conjunction with regular immunohistochemistry and fluorescence to automatically identify colon tissue types known as Spectral Histopathology (SPH).(Kallenbach-Thieltges et al. 2013) Good correlation between the techniques was found to ascertain the morphology of the tissues and several distinct structures and tissue types (lamina propria, muscularis mucosa etc.), and comparison made with traditional H & E reviewed by an expert pathologist. The accuracy of the Random Forrest statistical analysis method used was found to be around 95%.

Nallala et al. used FTIR spectroscopy to generate a new concept in analysis called spectral barcodes.(Nallala et al. 2013) IR images from 10 patients, cancer and normal, were clustered and classified according to their biochemistry to produce colour-coded images. These cluster images enabled the retrieval of specific IR signatures to differentiate cases of cancer and non-cancer epithelial components. Statistical analysis was performed on the specific signatures to identify the most discriminate wavenumbers from the two tissues and arranged in the form of spectral barcodes.

3.5 Infrared Spectroscopy Instrumentation and Set-up

FTIR spectroscopy uses the mid-infrared region of the EM spectrum and studies the interaction of infrared light with a sample. The typical pathway is described in Figure 3.3.

FTIR spectrometers use a light source where the light passes through an interferometer to provide an interferogram, which is the inverse Fourier transform of the spectrum. The interferometer allows the whole range of wavelengths of light to be measured at any one time, thus reducing the time taken to scan dramatically. Light is directed towards a specimen and absorbed at the frequency at which the incident radiation is equal to that molecule's vibrational mode. This loss is detected against the background at each frequency in the spectrum.

After light passes via a series of reflective optics through the specimen, the interferogram is scanned through using a moving mirror in the reference arm of the interferometer. The signal is measured with a single detector.(Kendall et al.

2009) The computer is then used to perform a Fourier transform on the data to convert it into a transmission spectrum (Barth 2007) (other modes are available such as reflectance, when reflected radiation from the specimen surface is measured). The ratio of the sample transmission spectrum to the background transmission spectrum provides the transmittance spectrum of the sample (removing the instrumental contributions to the spectrum).

A transmittance spectrum is the transmittance plotted against wavenumber. If all light passes through a sample then the transmission is 100% and absorbance zero and vice versa to varying degree. An absorbance spectrum is generated from logarithm to the base 10 of the reciprocal of the transmittance.
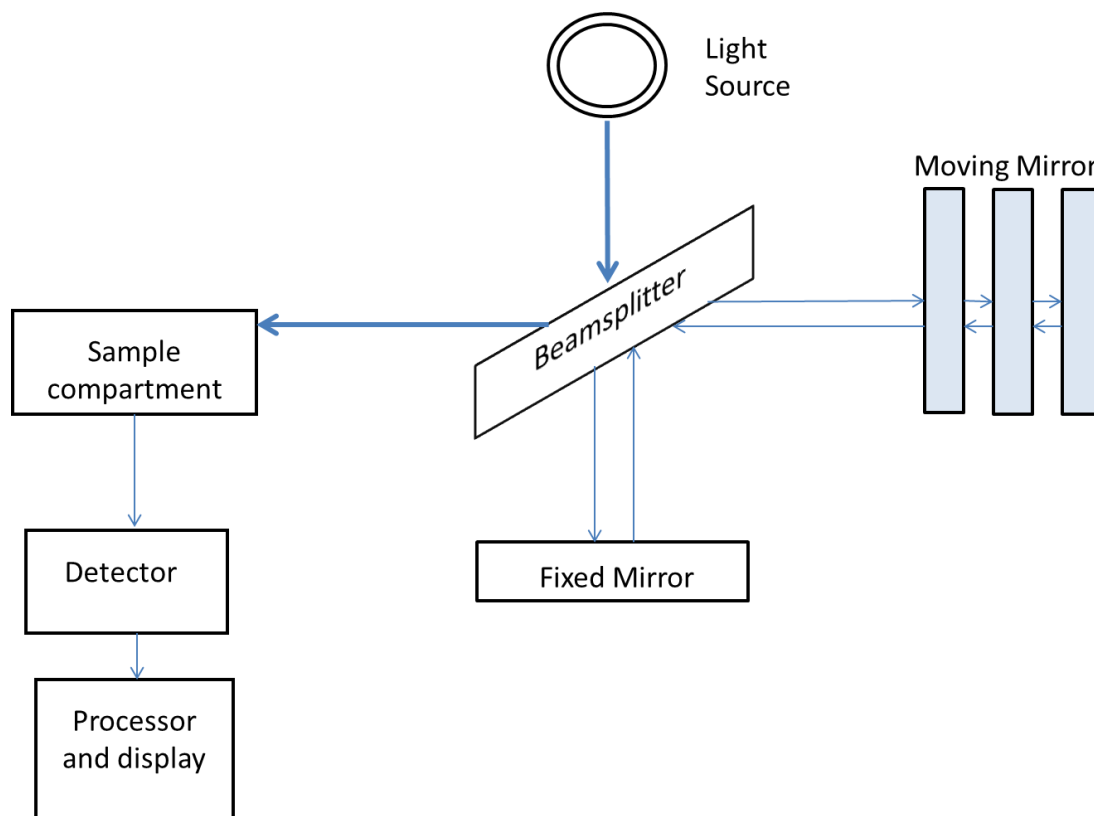
Figure 3.3: Schematic simplified representation of the basic components of an infrared spectrometer

Infrared systems have the following similarities; they are equipped with a microscope coupled to a light source, a computer and often have a sensitive liquid nitrogen cooled Focal Plane Array (FPA) detector, which if present

requires cooling to 77K. Liquid nitrogen is poured into a dewar for storage, and gradually the detector temperature reduces. This remains cool for a period of approximately 6 hours for both. Before any measurements are taken the user must ensure that the detector is cool enough and energy reading is appropriate.

3.6 High Resolution Infrared Spectroscopy

The University of Exeter houses one of only a few novel 'high resolution' spectrometers in the world. It is a prototype system developed by Agilent® with a high magnification optical system. The main differences between high and standard resolution are the optics- the high resolution optics have an additional magnifying element in combination with a high numerical aperture 15 x condenser. Together these give 40 x magnification (1.1 µm) compared with standard 15 x (5.5 µm) pixel size. This therefore has the potential to differentiate smaller features (potentially sub-organelle) and thus give greater amount of information per pixel which could increase the heterogeneity of data collected.

The understanding of this new technology is still in evolution, and as a consequence there is very little published data on this new technology. The University of Exeter team is one of a number of research centres involved in the Minerva® (MId- to NEaR infrared spectroscopy for improVed medical diAgnostics**)** project which is supported by the European Commission under the Seventh Framework Programme (FP7).(Napier 2012) This project aims to provide high resolution and tunable IR imaging so that specific spectral regions of interest can be selected to allow rapid imaging in the future. Consequently the set-up parameters of the experiments in this thesis, such as number of scans per sample, were therefore directed by the team based on as yet unpublished work from them. The first published high resolution data was in 2012 testing accuracy of the technique using brain tissue and Arctic Sea Ice Diatoms in comparison to standard resolution and Synchrotron light source. They found an increase in spatial and spectral detail, particularly in the fingerprint region, of comparable quality for signal to noise ratio in formalin fixed paraffin embedded tissue. Results were obtained using only a cheap bench top light source.(Findlay et al. 2015) 128 scans and a resolution of 4 cm^{-1} were used with samples mounted on barium fluoride slides, but only one case of

human tissue was analysed. A recent paper demonstrates the use of high-resolution spectroscopy at comparable magnification to the experiments in this thesis on both kidney and liver tissue, but focused on presentation and comparison with standard resolution rather than building a model for tissue classification as has been done in this thesis, and only a few pathological cases were actually analysed.(Sreedhar et al. 2015) Another team focused their experiments on the theory of light interaction at a reduced pixel size and calculated the smallest functioning aperture size to produce a pixel size of near 1.1 μm (128 scans at resolution of 8 cm^{-1}). Again this was presented in a format of comparison against the standard resolution using prostate tissue, rather than as a case series with analysis as in this thesis. They found that the smaller pixel size gave much more detail in the resulting images but at the expense of lower signal to noise ratio and proposed a computational method of noise reduction without increasing experiment time.(Reddy et al. 2013)

3.7 Thesis Objectives

The aim of this study is therefore to use high-resolution infrared spectroscopy to interrogate colonic tissue and analyse the resulting spectra. The benign pathological condition of epithelial misplacement and the difficulty associated with its diagnosis has been discussed at length in this thesis and it is here that I will focus my research as there is a clear clinical need for improvement in its diagnosis.

I will therefore attempt to use high resolution infrared spectroscopy to differentiate normal colonic epithelium from a range of common colonic pathologies that would be readily found at polypectomy and biopsy during BCSP colonoscopy, with particular interest in the use of infrared in the differentiation of cancer from epithelial misplacement.

CHAPTER 4

MATERIALS AND METHODOLOGY

4.1 Ethical Approval

Ethical approval for the project was granted by Gloucestershire Local Ethics Committee and was already in place at the beginning of the project. Those patients eligible for participation within the study were identified from historical hospital databases. No fresh tissue was collected for this particular project and so it was not necessary to consent the patients for tissue use.

4.2 Tissue Collection

Patients were identified from hospital databases where all pathology specimens are given a unique code from hospital records. Specific cases of epithelial misplacement and polyp cancer were identified by records kept by Professor N A Shepherd in his capacity as Regional and Local BCSP Pathology Lead.

Once the specific cases had been identified, the histology case number was identified from the patient number, and this was used to track the location of the tissue in storage. The historical cases had already been previously routinely processed and placed in formalin at time of excision and original investigations and then embedded by the Histopathology laboratory. All tissue had been kept in ideal conditions according to departmental protocol. Thus the paraffin embedded tissue blocks had to be identified for each case.

4.3 Histopathology

Routine processing had been carried out for all specimens used in this project. Historical samples from 2010 onwards were used (from the end of the last PhD student so that no cases were overlapping and all work was original).

Specimens were received by the histopathology laboratory fixed in formalin and examined macroscopically before being processed in cassettes and then put into hot paraffin wax to further fix and protect the tissue. Once the blocks had cooled and solidified they were then cut into 4 µm to 5 µm sections and placed

onto rectangular glass slides and undergo special H&E stain and immunohistochemistry to facilitate analysis under the microscope. All the cases had been reported by a Consultant Histopathologist and all abnormal results discussed at the Colorectal Multidisciplinary Team meeting (MDT).

Pathology cases collected for this project included the following:

1. Normal
2. Adenomas (low Grade dysplasia)
3. Cancers
4. Polyps with Epithelial Misplacement
5. Polyp Cancers

A secure database was designed to store patient details including hospital number, date of birth, time of biopsy, and the corresponding sample details such as research code and histopathology details. Each patient was assigned a unique identification code according to group.

4.3.1 Tissue Samples

100 cases were initially identified in total across the five pathology groups, 20 patients in each group. Historical samples were identified from June 2010 to May 2014. All these patients were managed according to the NHS protocols in place at the time. This research has not altered the management of any of the patients in any way. Some underwent surgery and further treatment at the time of their diagnosis and a few are still in follow-up programmes now.

Acquisition of specimens took place in the histopathology laboratory at Cheltenham General Hospital between April 2014 and June 2014. Once the tissue blocks had been identified and found, the blocks had to be further sectioned for the number of experiments necessary by laboratory technicians at Gloucestershire Cellular Pathology Laboratory.

The pathological diagnosis had already been made by one (or more – at this time, protocols were in place that assured that all diagnoses of cancer, whether in a polyp or otherwise, were confirmed by two consultant pathologists in Gloucestershire) Consultant as per routine processing protocols at the

Gloucestershire Cellular Pathology Laboratory in Cheltenham General Hospital. The reporting Consultant was not always the same person each time as cases are randomly allocated to Histopathology Consultants as part of their normal working day but each cancer diagnosis was confirmed independently by two consultant pathologists.

A list of historical patient diagnoses for every biopsy and tissue sample taken is kept on record by the Trust in the coding department. This information was accessed to identify the cases across the five groups initially. Once the list of cases had been identified, the H&E slide and block were retrieved for all. A single expert gastrointestinal Pathology Consultant (NAS) then reviewed all H&E slides to confirm the diagnosis made previously. Any cases with a questionable diagnosis or when the two consultants differed in opinion were excluded from the experimental protocol.

All cases of the cancer and epithelial misplacement (EM) had also been discussed at the local Hospital Colorectal multi-disciplinary team (MDT) meeting amongst a panel of expert Pathologists, Surgeons and Gastroenterologists, and management decisions made in advance of experiments being performed on remaining tissue.

All cases of dysplasia used were classified as low grade. All cancers were classified as at least T2/3, the polyp cancers all being T1.

Blocks were cut in contiguous sections of 5μm thickness and placed on slides as follows, see Figure 4.1:

➢ H&E (glass)
➢ IR slide (Calcium fluoride)
➢ H&E (glass)

Figure 4.1: Schematic representation of sections through the tissue block

An H&E was requested after every IR slide so that it could be ensured the pathology was visualised right through the 15µm thickness. This validated our sampling methodology and ensured that the pathology could still be found in latter sections. They were sectioned in continuity for the same reason which created a stack of contiguous sections. Every single slide was re-analysed by a specialist Consultant Histopathologist to ensure agreement with the initial diagnosis that had been recorded. Following analysis by the lead pathologist, several cases were deemed unsuitable or the pathology had been recorded incorrectly and so these were excluded.

67 cases in total were identified which met all the criteria. Their corresponding tissue blocks were then sent to the Histopathology laboratory for cutting according to the experimental protocol. This resulted in a selection of new stained H&E slides for each sample with corresponding colourless Calcium fluoride (CaF_2) slide sample (contiguous slides). On several occasions, more than one polyp was retrieved per patient, so the actual number of experimental cases was greater than total number of patients involved. A final total of 58 patients were included in the study, see Table 4.1.

Group	Normal	Dysplasia	Cancer	Polyp cancer	Epithelial misplacement
Patient total	12	10	8	13	15
Sample total	15	16	8	13	15

Table 4.1: Initial total number of patient and study samples for each of the five pathological groups

Calcium fluoride (CaF_2) was used as the infrared slide substrate because it does not give any signal which can interfere with the fingerprint region (1000-1800 cm^{-1} in infrared), see Figure 4.2 for the absorption spectra, whereas other substrates such as glass can prevent mid-IR transmission, since tissue absorptions below 2000 cm^{-1} are obscured by the opacity of glass. CaF_2 slides are comparatively expensive however, and great care was taken during handling to ensure that none were broken.
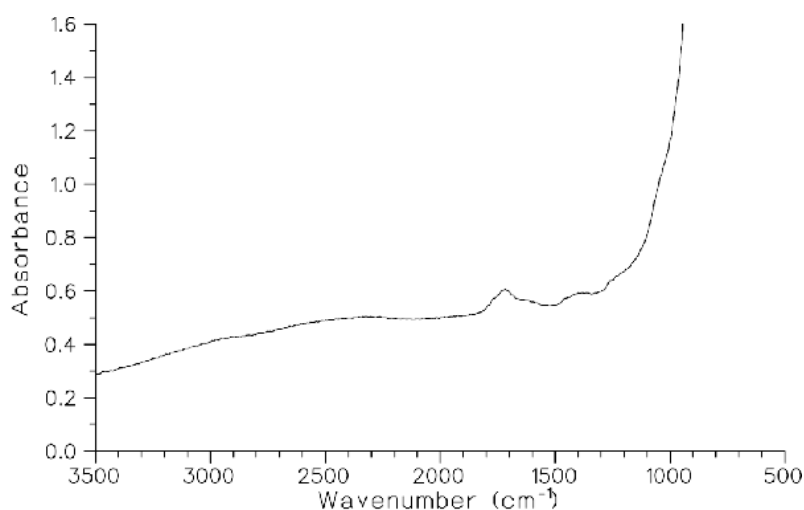
Figure 4.2: Calcium fluoride slide absorbance signal (Davis 2007)

The slides were all labelled and correlated with the specimen database so that a clear record of each sample and its history was kept. The slides and remaining blocks were able to be stored at room temperature under normal conditions. No special provision had to be made during storage or at the time of IR spectroscopic analysis. They were all stored together in pathology cardboard folders which protected them from scratching and other sources of physical damage however did not completely protect from microscopic dust and particles. The slides were not fixed in place so transport was difficult and all folders had to be kept upright at all times.

4.3.2 Image creation

For every experimental sample, a white light image of the corresponding H&E stained slide was carefully mapped, digitally recorded and transferred for printing on A4 paper. This allowed the specific pathology to be identified and recorded in colour- something that was not possible with just the CaF_2 slide. Because the slides were contiguous the H&E was accepted as a fair representation of the CaF_2 sample for comparison.
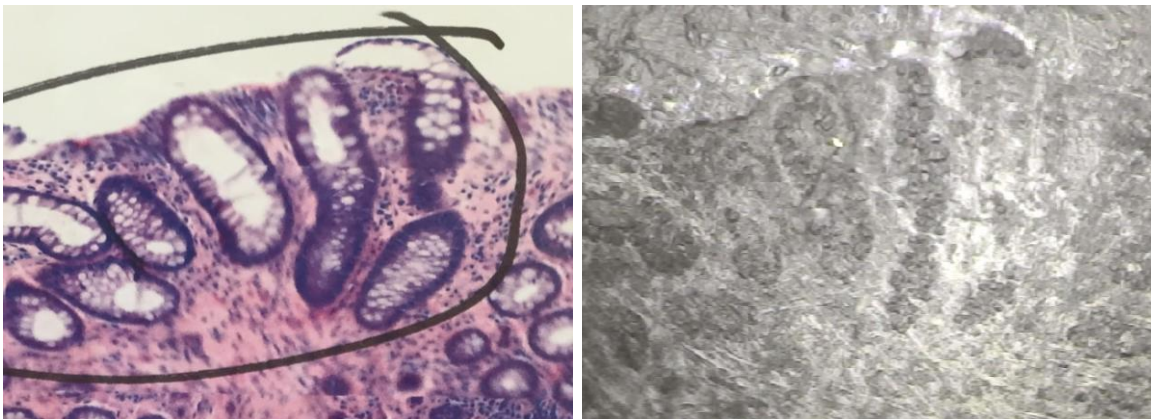
Figure 4.3: Histological H&E slide of normal glands (left), marked by pathologist with a circle, with corresponding white light image (right)

It was very important to accurately record the exact location of these glands within the tissue on the slide so that they could be re-located at a later date for

infrared analysis. For each sample therefore, a hand drawn map of exact area and location was recorded, using structural tissue landmarks and local features to use at a later date.

The expert pathologist was then able to manually select the exact glands required which ensured the quality of the selected glands. In many samples there was evidence of a heterogeneous population of glands within one specimen, and more than one pathology was noted to be present across the sample, for example the presence of both normal and dysplastic glands next to each other on one slide coded as dysplasia. The ability to identify and record the exact glands enabled the reviewing pathologist to be very precise. It was at this point that it was realised that often there was more than polyp section on the H&E slide, and this did not always correspond with the specimen on the CaF_2 slide, so a number had to be reprocessed to ensure that the glands could be identified on both the H&E and corresponding CaF_2 slide.

These reference maps were invaluable when it came to switching between the colourful stained slides to colourless unstained ones, and were particularly useful in areas with difficult merging glands where discrimination of one area from another was very difficult under the microscope.

4.4 Infrared Measurements

Experiments were conducted across the dataset over a period of 8 months from June 2014 to February 2015 and not in group order, to ensure that results were not skewed by experimental conditions on a particular day.

Infrared experiments were conducted across two sites, initially at the Biophotonics Department at Gloucestershire Royal Hospital where a small number of cases were analysed at standard magnification and then at high magnification (see Section 3.6) in the Physics Department at the University of Exeter, Streatham Campus, between June 2014 and February 2015. Please see Appendix 1 for preliminary work completed on a minimal dataset on the standard resolution FTIR spectrometer at the Biophotonics Department in Gloucester. All experiments were conducted in purpose built rooms and in

accordance with safety guidelines with local risk assessment documents in place.

Initial measurements were taken in Leadon House in the Biophotonics Department at Gloucestershire Royal Hospital, Gloucester. Two measurements per pathology group were taken in Gloucester to standardise and perfect the protocol. All other measurements for all samples were taken at the University of Exeter, Exeter in the Physics Department. Infrared map measurements were collected in Gloucester using a PerkinElmer® Spectrum One (Spotlight) 400 Fourier transform infrared (FTIR) spectrometer (PE system) and at Exeter using an Agilent® 620 FTIR microscope coupled to 670 Agilent spectrometer both in transmission mode.

The PE system provides 6:1 imaging, resulting in a pixel size of 6.25µm. A charge couple device (CCD) camera under white light LED illumination attached to the microscope and computerized x-y-z stage enable the specimen to be visualised and to capture a visible white light image. When the system is started up it checks that each component (stage controller, computer, microscope components, and Spectrum One) is attached and communicating well. Problems are highlighted on the panel of Spectrum One (each component must have a green light to start). The detector moves in a raster formation and collects 16 pixels at a time, see Figure 4.4.
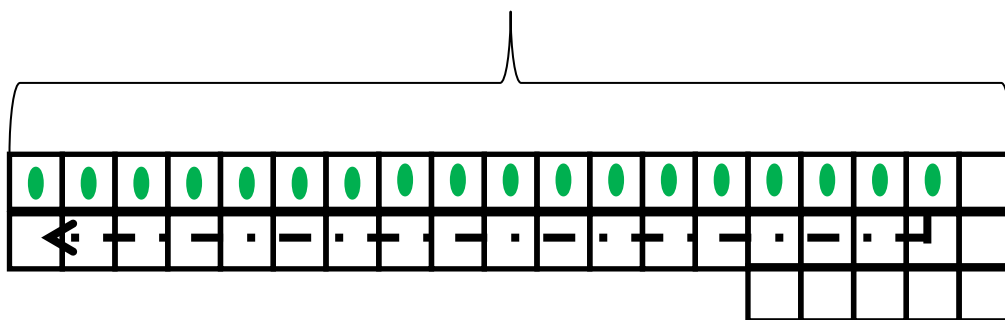
Figure 4.4: Schematic representation of Perkin-Elmer® detector raster movement. Each coloured dot represents the acquisition of a pixel (16 in total)

In contrast to the PE system, the Agilent, when set to high-resolution mode, has a pixel size of 1.1µm and the 128 x 128 FPA detector moves in a snake-like formation collecting one full spectral image of information at a time (see Figure 4.5). The Agilent has more manual set-up tasks at the beginning of each session which includes focusing the light beam by 'centerburst' and verifying the energy. These are carried out within the spectrometer.
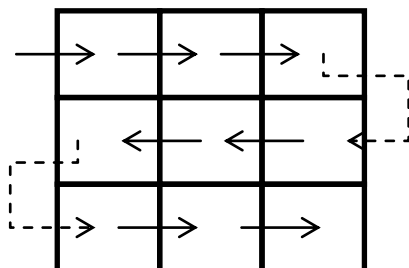
Figure 4.5: Schematic representation of Agilent® detector snake-like movement

4.4.1 High magnification infrared spectra

All 67 specimens were analysed. Experiments were performed at the University of Exeter Physics Department on the Agilent Infrared spectrometer at high magnification. This is one of only a few high resolution spectrometers in the world currently and as such the experimental design evolved from a combination of recent unpublished work the team in Exeter had done before my arrival, and standard recognised infrared experimental protocol.

A high magnification background takes approximately 20 minutes to acquire at a spectral resolution of 4 cm^{-1} with 256 scans, which is in total approximately 30 minutes including set up time. A corresponding background was initially taken for every single sample, however it was soon realised that in some samples the large polyp specimen took up the whole surface area of the small CaF_2 slide and as a consequence there was no space on the CaF_2 slide free from paraffin or sample to take a background from. After a few experimental changes with background type and background: CaF_2 sample ratio, it was decided that a separate blank CaF_2 slide would be used and background performed every third sample to increase time efficiency. This seemed to give a reasonable signal to

noise ratio with good data. If more than three samples in a row were attempted then the signal to noise ratio would decrease and intensity map quality were noticeably poorer. The liquid nitrogen supplies were topped up before every background and subsequent sample analyses were performed. Measurements were therefore standardised as much as possible for temperature and light source average light intensity (maximum light intensity to integration time: part manual, part automated set up process for each new background).

Experiments were conducted using the linear array detector of the Agilent 620 FTIR microscope coupled to 670 Agilent Spectrometer in transmission mode using a pixel resolution of 1.1 μm and a spectral resolution of 4 cm^{-1}. 64 scans per pixel were performed for the samples. Spectral information was acquired from mapping in a raster movement across 3x3 tiles. 128x128 pixels make up one tile and the data for each is acquired simultaneously. Each pixel is 1.1 μm in diameter, so a tile is 140 μm diameter; therefore each area of colon mapped was (3x140) 420 μm x 420 μm in diameter. Because data for each tile is acquired simultaneously, the time taken for a map of this size is much quicker than for standard resolution previously acquired on the PE system; however the overall mapped area is generally smaller (depending on area chosen). In most samples 3x3 tile size map enabled information to be gathered on one or two colonic glands only. A 3x3 tile takes approximately 45 minutes to map. Therefore each cycle of background to 3 sample slides took a minimum of 2 hours 45 minutes for spectral acquisition alone.

All samples were processed once. On a few occasions it was clear that the signal to noise ratio of the spectra on the raw data was poor and therefore the samples were repeated according to protocol.

4.5 Data Analysis

The spectrometer generates an inferogram based on all the biochemical information from the sample analysed. The output of spectral mapping is known as a hyperspectral image or hypercube. These are both discussed in three dimensions; the X-Y plane gives the spatial dimension (wavenumber and intensity) and the Z "3rd dimension" which represents the complete spectrum per pixel, which reflects the chemical composition of the sample.(Romeo et al.

Page | 71

2008) There is a huge amount of unrefined information which needs to be analysed and processed to understand its significance. The huge volume of data can prove challenging to interpret qualitatively due to the subtle change in intensity and shift in peaks that might be noted. Thus the field of chemometric, multivariate statistical analysis, was developed with various models to analyse the data.

MATLAB® computer programmes were used for all analysis. The hypercube for each sample is loaded into MATLAB® for data analysis, the scripts for which were written in-house by an expert in chemometrics.

4.5.1 Spectral Pre-Processing

The generated interferogram was automatically Fourier-transformed and a transmission spectrum and absorbance spectrum subsequently calculated:

$$A = log_{10} \frac{1}{T} = -log_{10}T = -log_{10}\frac{I}{I_0}$$

A= Absorbance, *T*= Transmittance

$\frac{I}{I_0}$ = Ratio of radiant fluxes

Absorbance data was vector normalised to remove differences in the spectra due to changes in measured absorbance due to sample thickness. Certain wavenumbers were excluded- those which were outside the fingerprint region, 1800- 900cm^{-1}. Outside of these wavenumbers the mean spectra were found to be very noisy and therefore not useful because of reduced signal-to-noise ratio.

Following those steps, the wavenumber area under the largest peak (1600-1700 cm^{-1}) was used to visualise each image in an Intensity Map. The first pre-processing step was to select the specific spectra from only the epithelium as area of interest as these spectra were predicted to be the most discriminatory when compared across groups. The raw data of each sample was uploaded into MATLAB® and epithelium selected from the intensity map following the pre-

processing steps. The epithelial spectra were then selectively saved and used for comparative analysis, an example of which is shown in Figure 4.6 below. All other spectra were discarded from analysis.

Regions of interest were then exported for chemometric analysis.
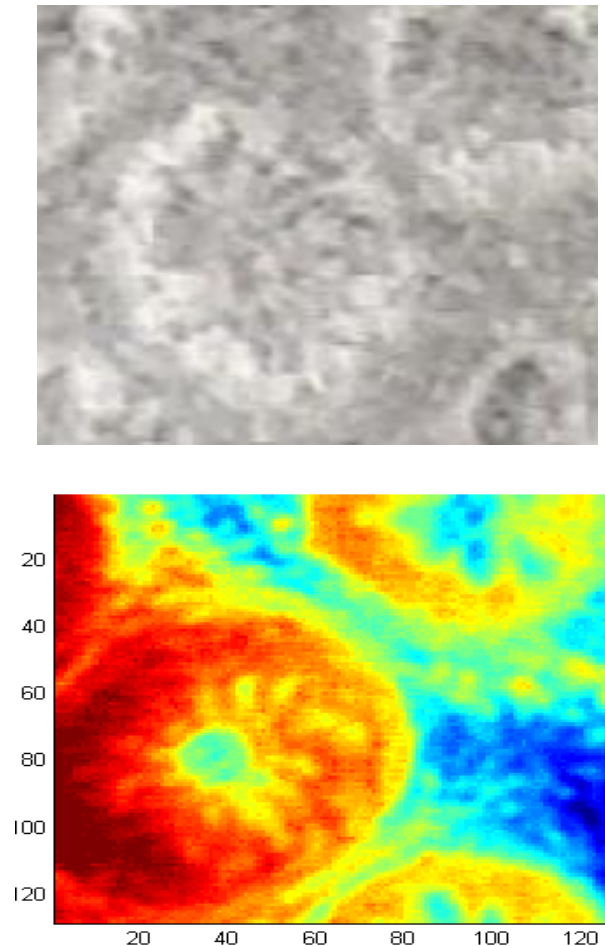
Figure 4.6: (Above) White light image of a cluster of normal glands with corresponding intensity map of the same region (Below). In this example the epithelium can be clearly identified

A paraffin correction was developed and built into the MATLAB® scripts to counteract the influence the paraffin peak had on the chemometric analysis. This was essentially used to negate the presence of paraffin in the mean spectra and to reduce the effect the paraffin would have on separating out the groups by principal component analysis. This was based on an Extended

Multiplicative Signal Correction (EMSC) to reduce variation in the paraffin signal, as described by Nallala *et al.*(Nallala et al. 2013)

4.5.2 Chemometric Analysis

Principal Component Analysis (PCA) is an unsupervised technique first applied to reduce data size by finding the largest variance within the dataset. It is a transformational tool used to make a predictive model based on multivariate analyses and assumes no prior knowledge about the dataset. It can reveal the data as a description based on variance as a set of co-ordinates in a data space. The analysis generates principal components, (PC) '1' being the largest measure of variance, PC2 the next greatest variance and so on. The further down the list of PCs the greater the probability that it is just noise rather than real spectral differences that they describe. Only the first few components are used to produce a multi-dimensional figure representative of the transformed data (so the figure demonstrates the first few PCs of greatest significance). The maximum number of PCs is either the number of rows (spectra) or the number of columns (wavenumbers) in the data set, whichever is smaller.

Analysis Of Variance (ANOVA) is then applied to find which PCs demonstrate the greatest significance in difference between groups i.e. PCA scores and use the most discriminative ones.

Linear Discriminant Analysis (LDA) then uses the selected PCs and Gold Standard pathology information to train the classification model by maximising the separation between the groups and reducing variance within groups using a linear function. This is a supervised classification technique.

Ideally the final step is to independently 'test' the classification model developed in the steps described above with new tissue samples, to see if the model can correctly identify the tissue samples. However, due to limited resources (time and financial) the final step to test the model was with a Leave One Sample Out Cross Validation (LOSOCV) method. This effectively tests the model with a full set of independent or brand new patients. All of the data from a single test set is removed from the model and the rest of the data is then re-run to generate the training set. The test sample is then predicted by the model to determine

accuracy. This is repeated with each sample being left out in turn (not necessarily leaving out a single patient as a few cases may have come from a single patient in the pathology group) and a contingency table generated which demonstrates the sensitivity (true positives correctly identified) and specificity (true negatives correctly identified) of the method.

4.5.3 Peak Assignments

Spectral peak assignments are determined in order to identify significant biochemical differences between different tissue types. Spectral peaks can be identified with regards to their specific biochemical properties and is carried out using previously published data. There is often significant overlap between peaks and so true identification of the biochemistry of a sample can be complex. Recent peak assignments, described by Movasaghi et al,(Movasaghi et al. 2008) were used for the experimental classification, details of which will be discussed further in chapter 5.

CHAPTER 5

RESULTS

All data from the 67 cases analysed were transferred from Exeter to be processed in Gloucester using MATLAB software with in-house tailored programming. In total over 2 million spectra for all groups were generated for analysis.

Two cases were immediately excluded from the analysis (in the dysplastic and polyp cancer groups) because the data was corrupted and was unable to load in MATLAB and therefore unusable. This gave a total of 65 samples for analysis, see Table 5.1.

Group	Normal	Dysplasia	Cancer	Polyp cancer	Epithelial misplacement
Sample Total	15	15	8	12	15

Table 5.1: Actual sample numbers per group

The spectral data for all samples were loaded as a hypercube into MATLAB and data pre-processing applied as discussed in Chapter 4.

The intensity map (of 1600-1700 cm^{-1} specifically) was compared to the corresponding white light image and the H&E to identify the detail of the sample areas and particular spectra in areas of interest could subsequently be manually selected.

The selected spectra representative of the epithelial area for each sample were separated into the five pathological groups. The mean spectra for all samples within each group were then plotted on a graph together, offset for clarity.
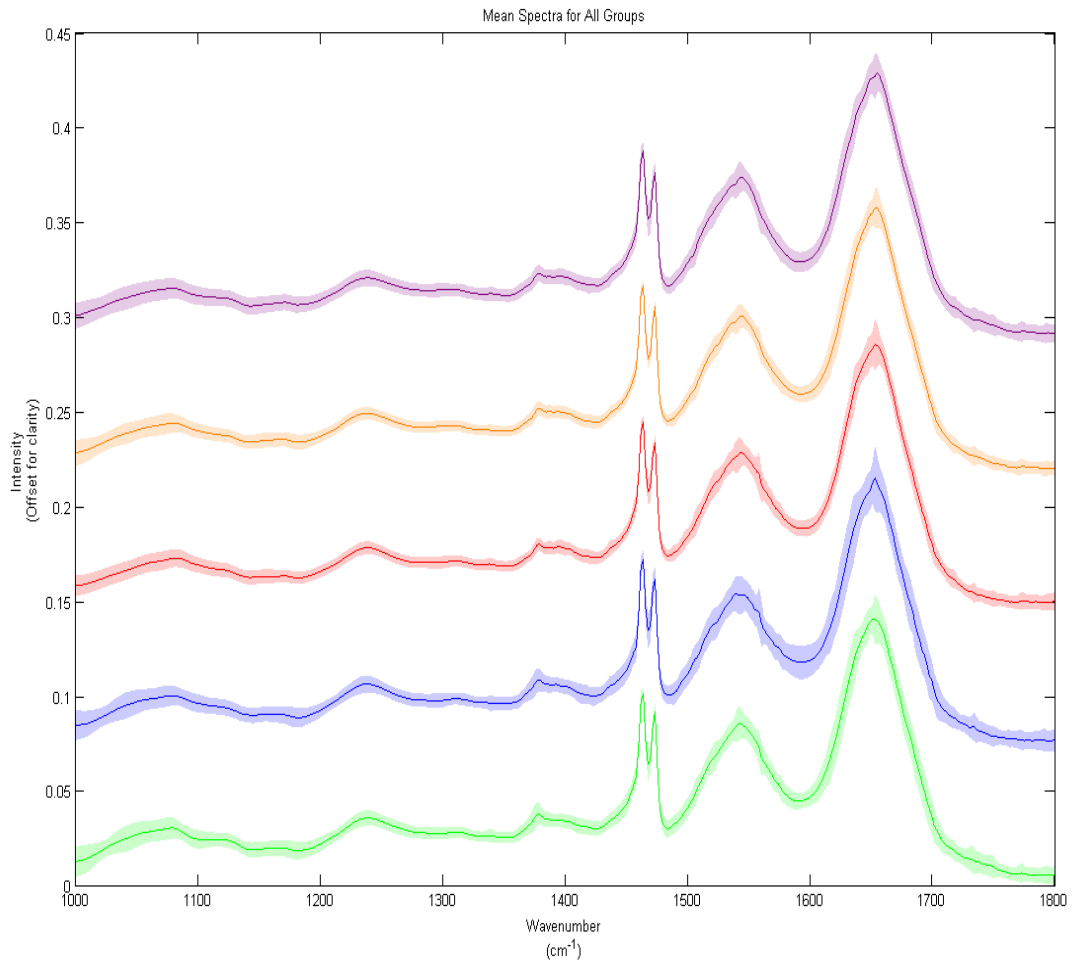
Figure 5.1: Mean spectra offset for all groups plotted with standard deviation.

Green= Normal, Blue= Dysplasia, Red= Cancer, Orange= Polyp Cancer, Purple= Epithelial Misplacement

The standard deviation (SD) +- 2 of the mean spectra per sample was calculated for each group (see Figures 5.2-5.6).
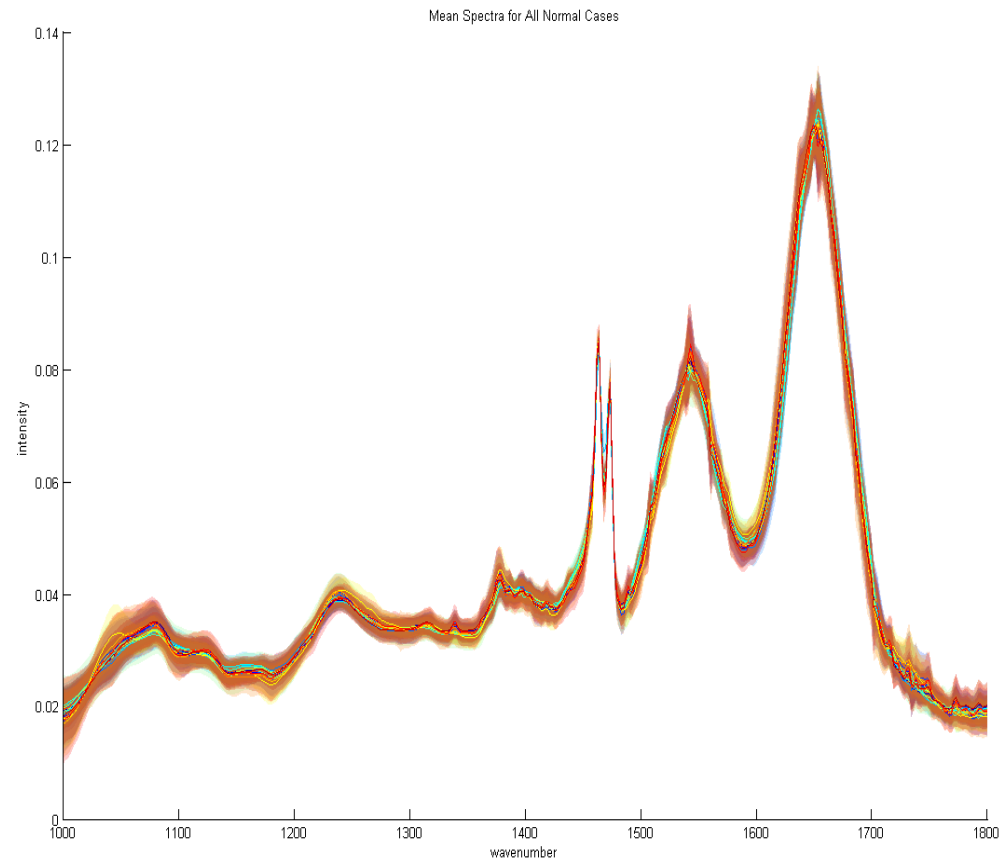
Figure 5.2: Mean spectra with SD for all cases plotted together for Normal group
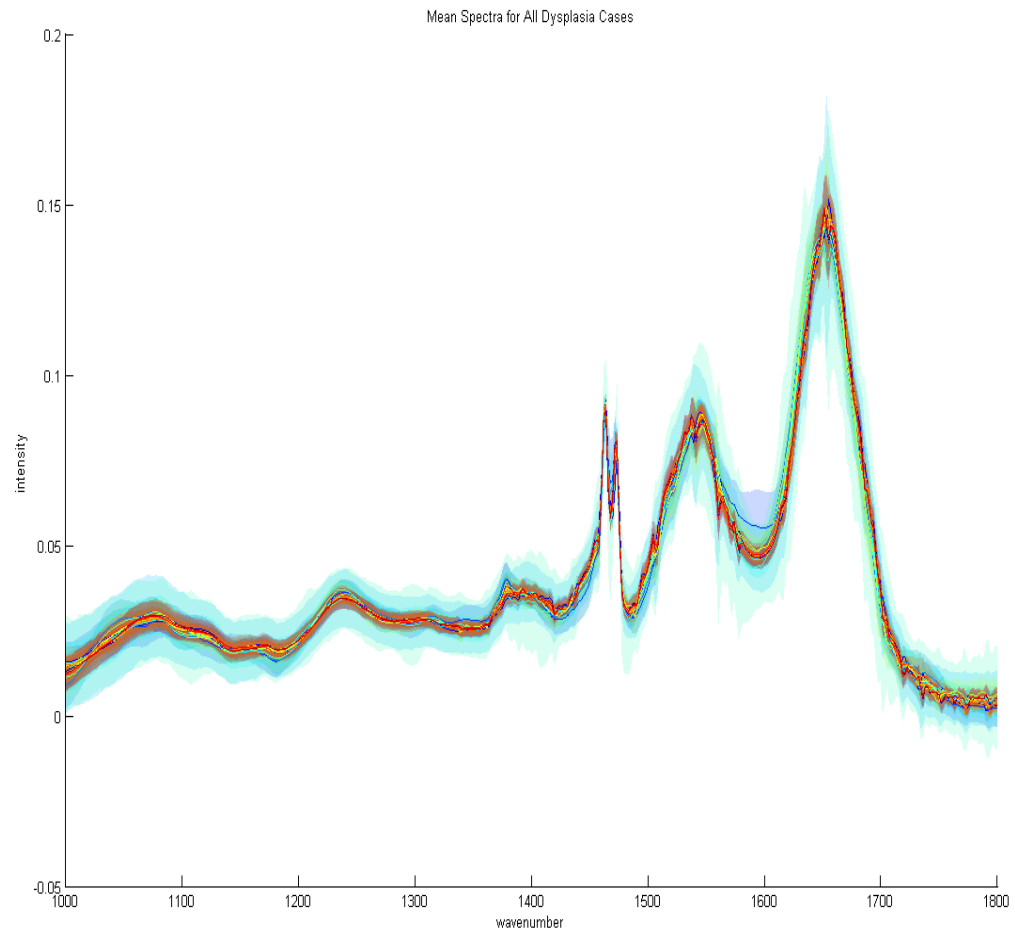
Figure 5.3: Mean spectra with SD for all cases plotted together for Dysplasia group
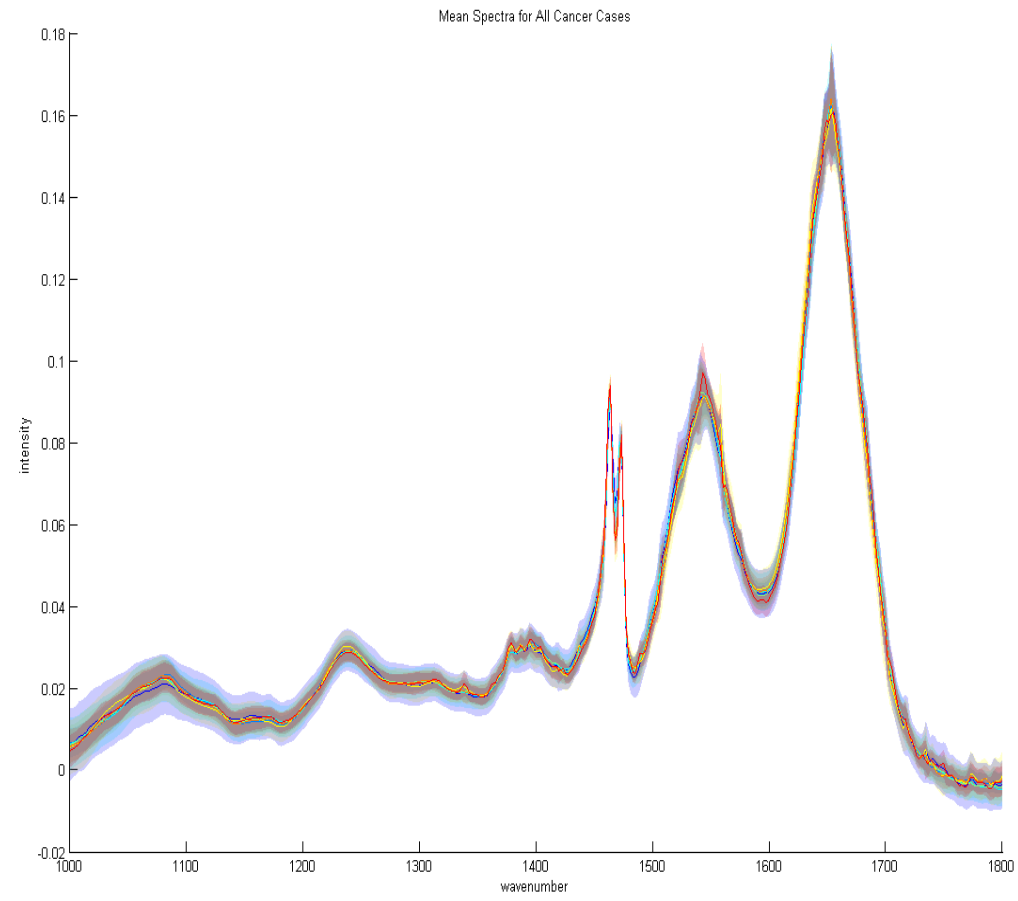
Figure 5.4: Mean spectra with SD for all cases plotted together for Cancer group

Mean Spectra for All Polyp Cancer Cases
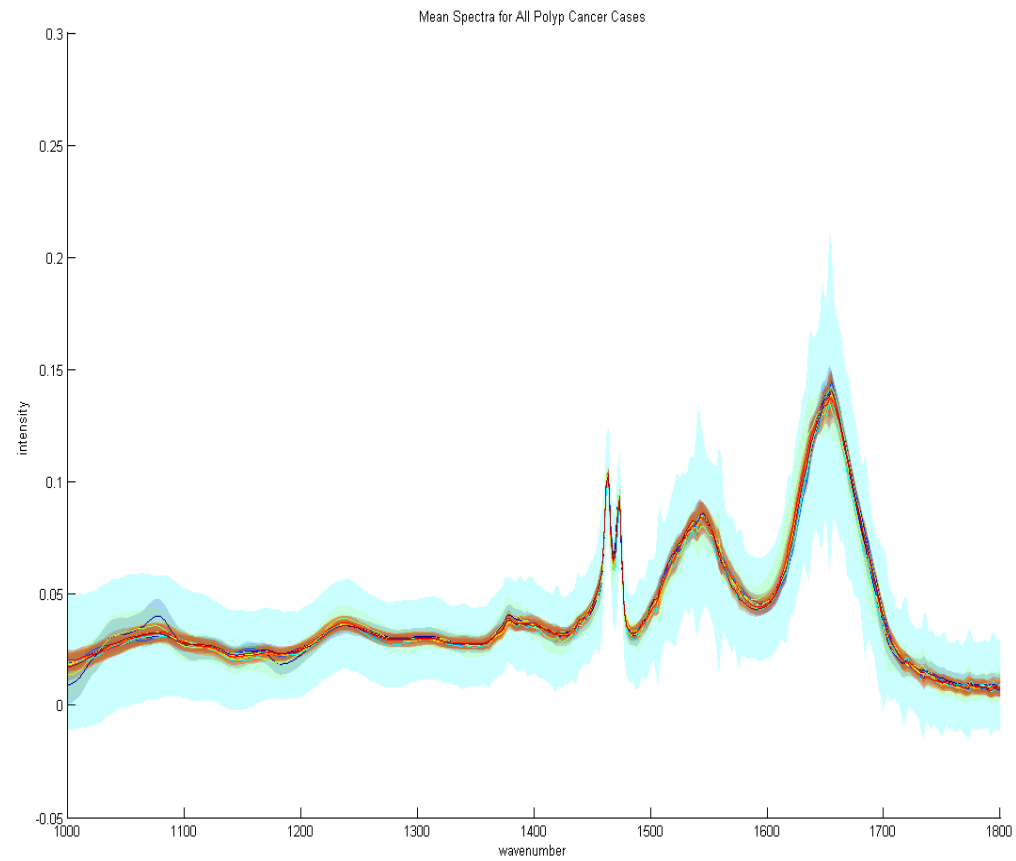
Figure 5.5: Mean spectra with SD for all cases plotted together for Polyp Cancer group
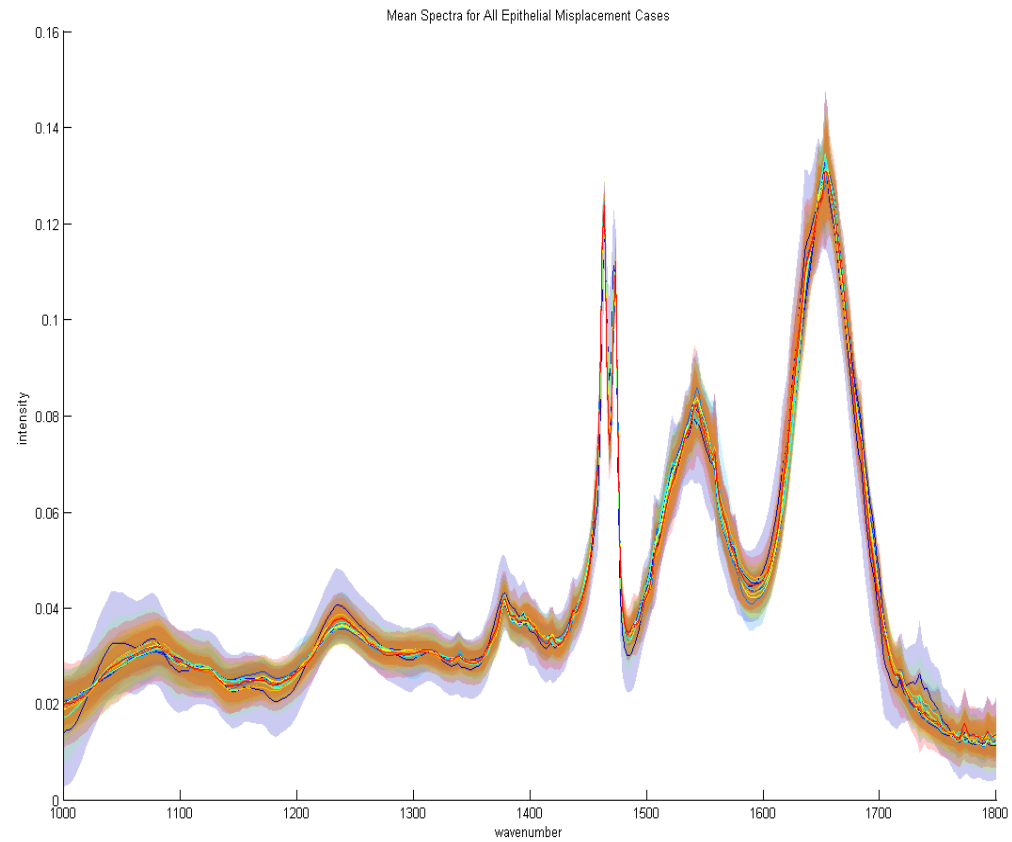
Figure 5.6: Mean spectra with SD for all cases plotted together for Epithelial Misplacement group

Manual assessment was made of the quality of the mean spectra per sample and five samples were identified as having an unusually wide SD (out-of-keeping with the rest of their group). These cases correlated with experimental difficulties that had been recorded at the time of experimentation:

- On three samples taken on the same day, B2_1, C1, E1 an annotation had been written "poor quality, high noise to signal ratio, obvious problems". Shortly after the Agilent technician visited Exeter to fine tune the Agilent Spectrometer.
- On sample B3_1 an annotation had been written "air conditioning broken in room" and was the last sample of the day. These factors coupled together was a valid reason to assume that the overall temperature had warmed up and therefore affected the quality of spectral data in the last sample
- On sample D14 an annotation had been written "problem with background. Had to use previous" so it would have been the fourth sample for the one background which may have affected the quality of spectral data

Following manual exclusion of cases the mean and SD were then re-plotted for each group (See Appendix A, Figures 1.1-1.4; no exclusions were made from the Normal group so this was not re-plotted).

It was decided that there was therefore adequate justification for an automated process to identify samples with poor quality spectra and replicate the manual process without adding any selection bias. This would enable poor quality mean spectra to be extracted from the dataset, but accepted that this might also exclude some that had not been identified manually and might therefore have been of 'acceptable' quality. Optimal values were chosen empirically to balance extraction of poor spectra without risk of reducing total case number too much by extracting reasonable spectra, and the following was set for all; sample exclusion made if the SD threshold is greater than 0.006 of 2 x SDs for greater than 60% of all selected wavenumbers.

This selection method excluded the following samples; B2-1, B3_1, D14 and E1 which was in-keeping with those that had been identified as part of the manual selection process, although not all. These 4 samples were then excluded from

any further analysis. The mean spectra with SD were then re-plotted (see Figures 5.7-5.10). Again no exclusions were made from the Normal group so that remains unchanged from the first graph plotted (Figure 5.2).

The total number of spectra for all groups was 1,706,027. The group totals following epithelial selection and then extraction of poor quality spectra were as follows see Table 5.2.

GROUP	TOTAL SPECTRA
Normal	178,166
Dysplasia	479,595
Cancer	341,445
Polyp Cancer	306,381
Epithelial Misplacement	400,440

Table 5.2: Total number of spectra per group

Mean Spectra for All Dysplasia Cases Post Automated Exclusion
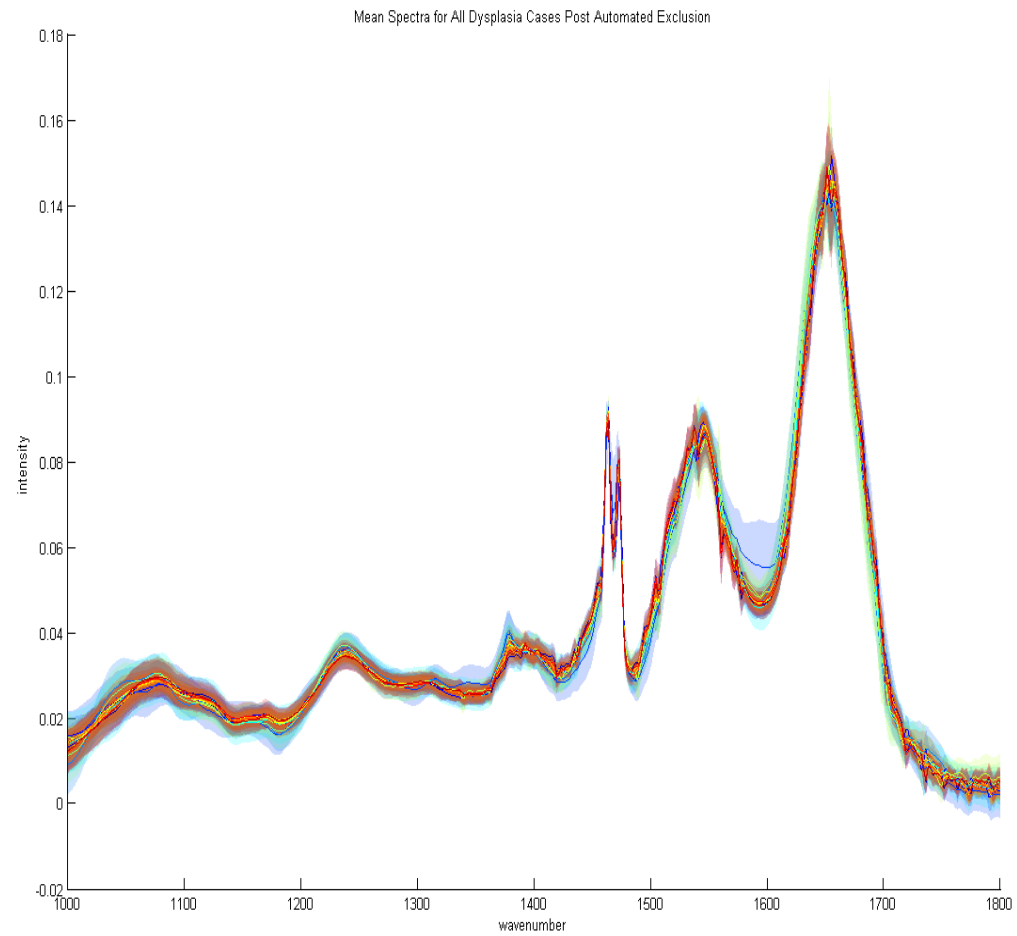
Figure 5.7: Mean spectra for all cases plotted together for Dysplasia group post automated exclusion
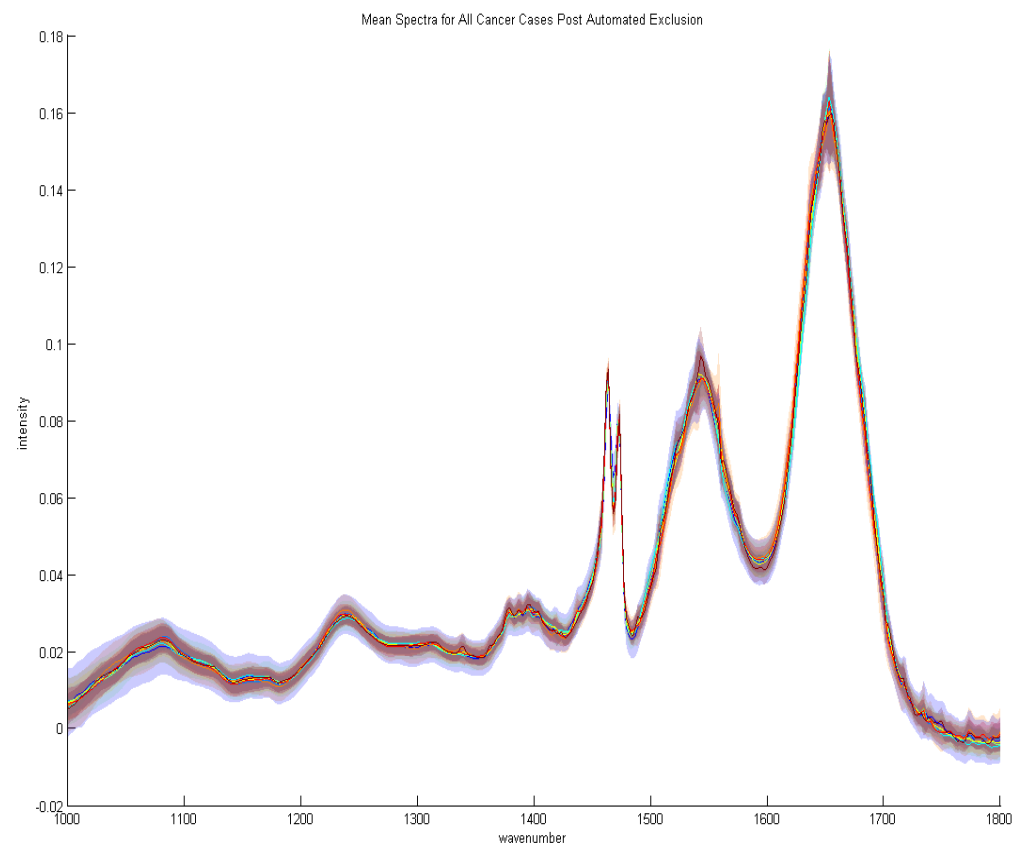
Figure 5.8: Mean spectra for all cases plotted together for Cancer group post automated exclusion

Mean Spectra for All Polyp Cancers Post Automated Exclusion
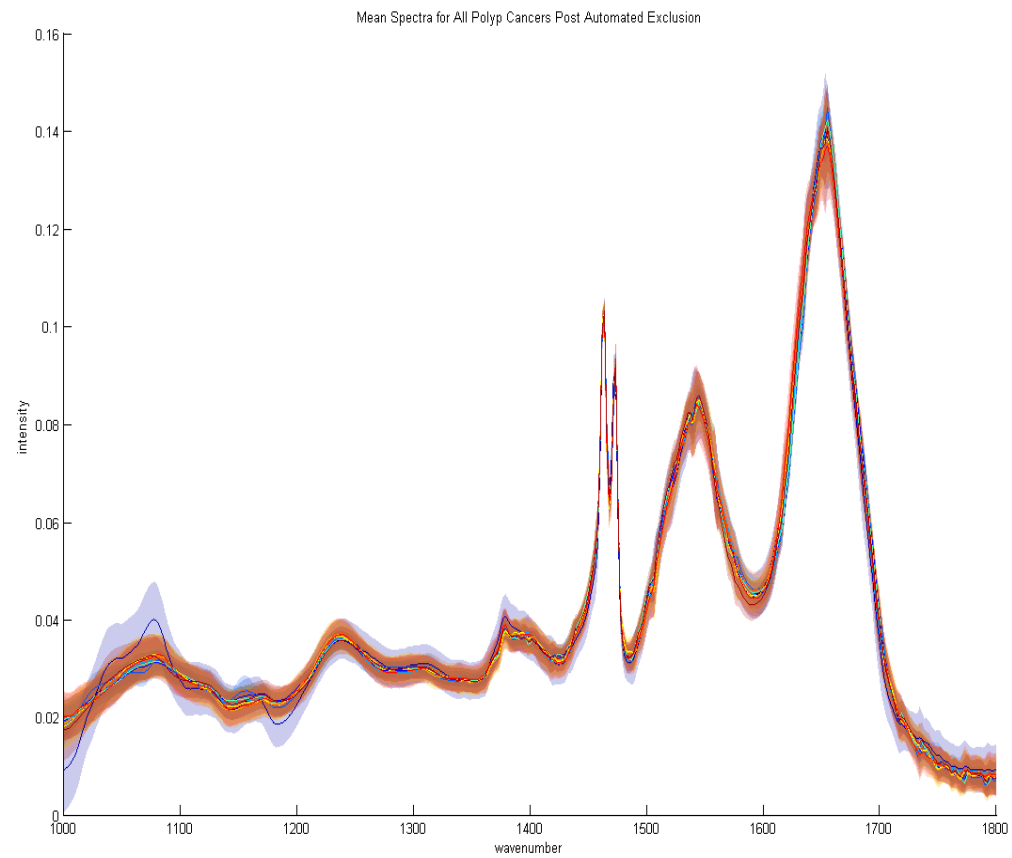
Figure 5.9: Mean spectra for all cases plotted together for Polyp Cancer group post automated exclusion
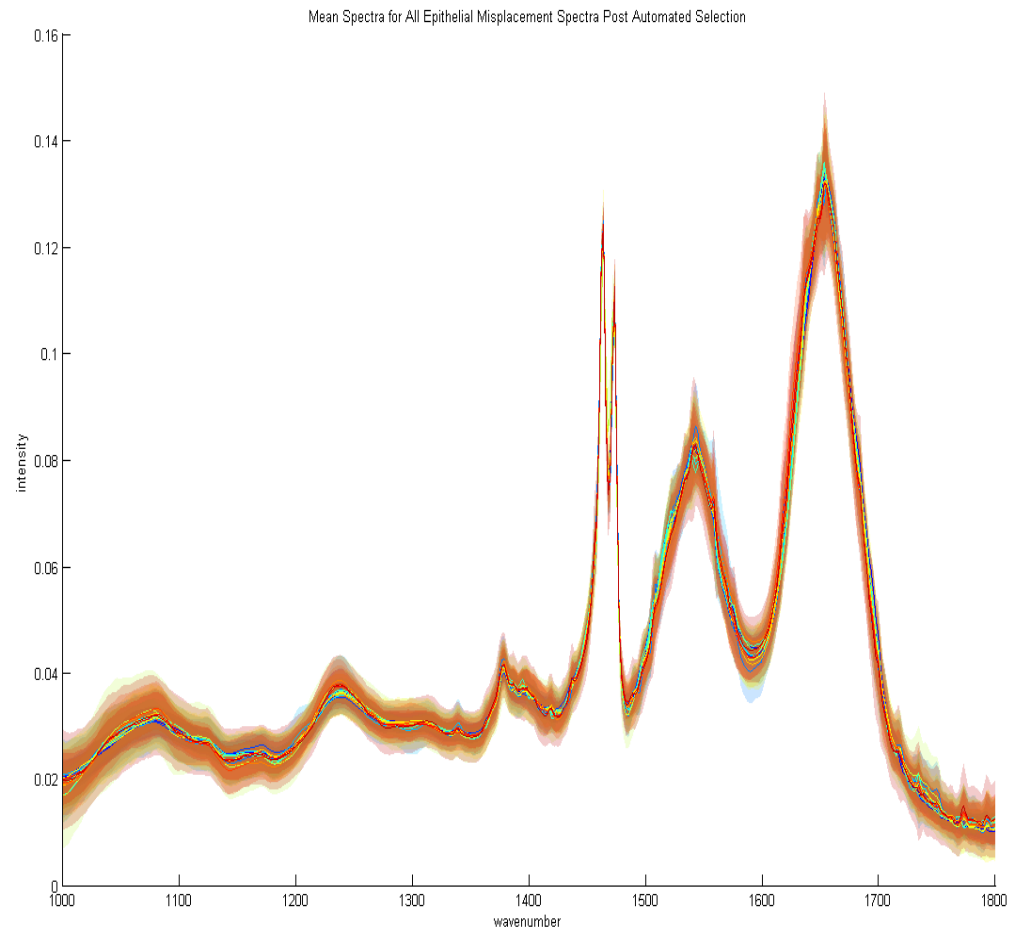
Figure 5.10: Mean spectra for all cases plotted together for Epithelial Misplacement group post automated exclusion

5.2 Univariate Analysis

Figure 5.11 shows the mean spectra as previously described in Figure 5.1, however values of the infrared peaks that demonstrate change between pathology groups have been identified and annotated. These correspond to the specific molecular biochemistry present in the tissues.
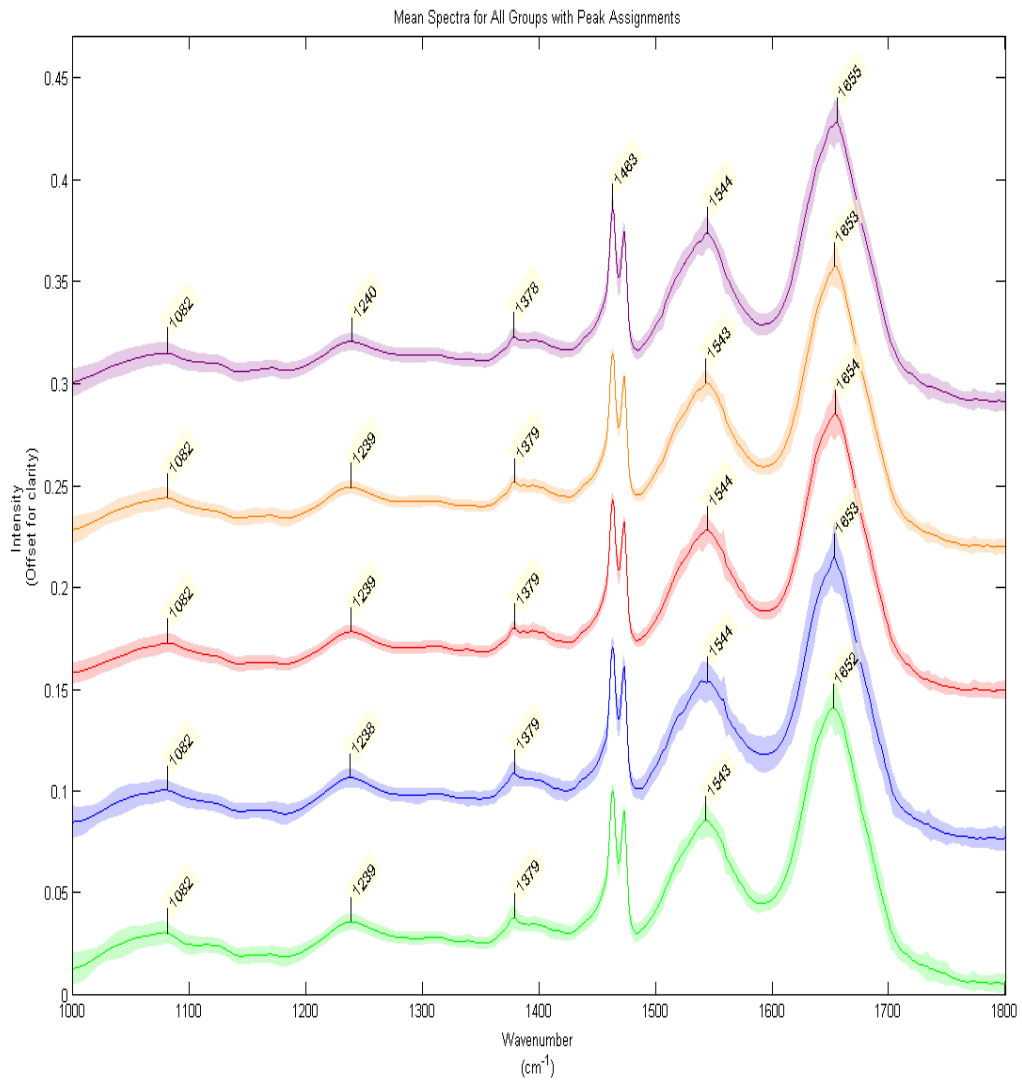
Figure 5.11: Mean spectra offset for all groups plotted with standard deviation and peak assignments labelled.

Green= Normal, Blue= Dysplasia, Red= Cancer, Orange= Polyp Cancer, Purple= Epithelial Misplacement

Table 5.3 demonstrates possible bonds responsible for these peaks.(Movasaghi et al. 2008) This is well corroborated by biochemical standards assigned courtesy of in-house Biophotonic records.

IR Peak (cm^{-1})	Nearest quoted (cm^{-1})	Assigned Bond	Tissue/substance	Author
1082	1082	νPO_2^- Symmetric band of the normal cells Phosphate band Collagen	Oral Tissue Colon Lymphoid tissue	(Fukuyama et al. 1999) (Rigas et al. 1990) (Andrus and Strickland 1998)
1239	1238/39	PO_2^- Asymmetric stretching of nucleic acids & phospholipid Amide III Triglyceride	Oral tissue	(Fukuyama et al. 1999)
1379	1380	δCH_3 of lipid, protein, triglyceride C-O Stretching C-H Deformation N-H deformation	Prostate, Colon Brain	(Paluszkiewicz and Kwiatek 2001; Richter et al. 2002) (Dovbeshko et al. 1997)
1483	1482/3/5	C8-H coupled with a ring vibration of guanine	Brain	(Dovbeshko et al. 1997)
1544	1544	Amide II bands C-N stretching & CHN bending vibrations	Fibroblasts	(Wood et al. 1996; Huleihel et al. 2002)

<u>1652</u>	1652	Amide I	Cervix	(Mordechai et al. 2004)
<u>1653</u>	1653/4	C2==O cytosine C==O, C==N, N-H adenine, thymine, guanine, cytosine (DNA)	Brain 'Tumour cells'	(Dovbeshko et al. 1997) (Dovbeshko et al. 2002)
<u>1654</u>	As above			
<u>1655</u>	1655	Amide I (α-helix protein conformation) Amide I (ν C==O, δ C-N, δN-H) C==O cytosine C==O, C==N, N-H of adenine, thymine, guanine, cytosine Peak of nucleic acids due to the base carbonyl stretching and ring breathing mode Amide I has some overlapping with the carbonyl stretching modes of nucleic acid	Brain, Prostate Brain 'Tumour cells' Brain Cervix Cervix	(Yoshida et al. 1997; Paluszkiewicz and Kwiatek 2001) (Dovbeshko et al. 1997) (Dovbeshko et al. 2002) (Dovbeshko et al. 1997) (Chiriboga et al. 1998) (Fung et al. 1996)

Table 5.3: Peak assignments from the literature

5.2.1 Derivative Spectra (Second Order)

Derivation is used to eliminate random fluctuations in the baseline (first derivative) and slope (second derivative) of the absorption spectra. Calculating derivative of a mean spectrum is a method of identifying peak frequencies of biochemical components and therefore provides qualitative and quantitative information. It also causes effective narrowing of spectral features, which may enhance specificity of the method.

Representative spectra of the five pathology groups in absorbance units and in first and second derivative are shown in Figures 5.12-5.16 were calculated using an in-house MATLAB® programme. The peak assignments have again been annotated on each group, the positive peaks demonstrating the exact wavenumber of maximum absorbance of the peak. No particular shifts in frequency (>10 wavenumbers) were observed over the entire region (1800-900cm^{-1}) in any of the groups. The comparison of group spectra showed several differences in the peaks present, all 2nd derivative spectra are multiplied by -1 for the Figures.
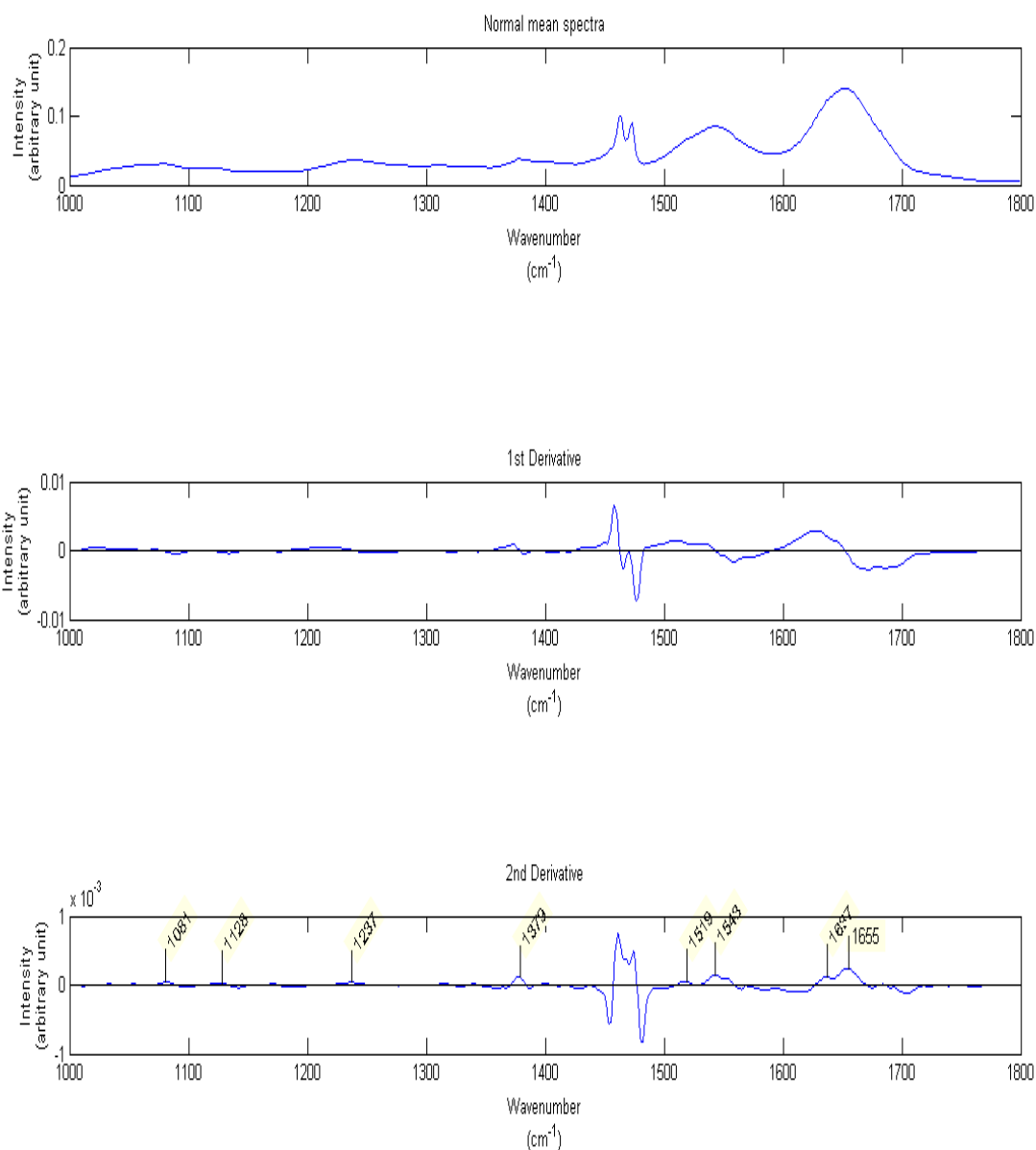
Figure 5.12: Normal group derivatives; (Top) Mean spectra, (Middle) 1ˢᵗ order derivative, and (Bottom) 2ⁿᵈ order derivative for Normal group
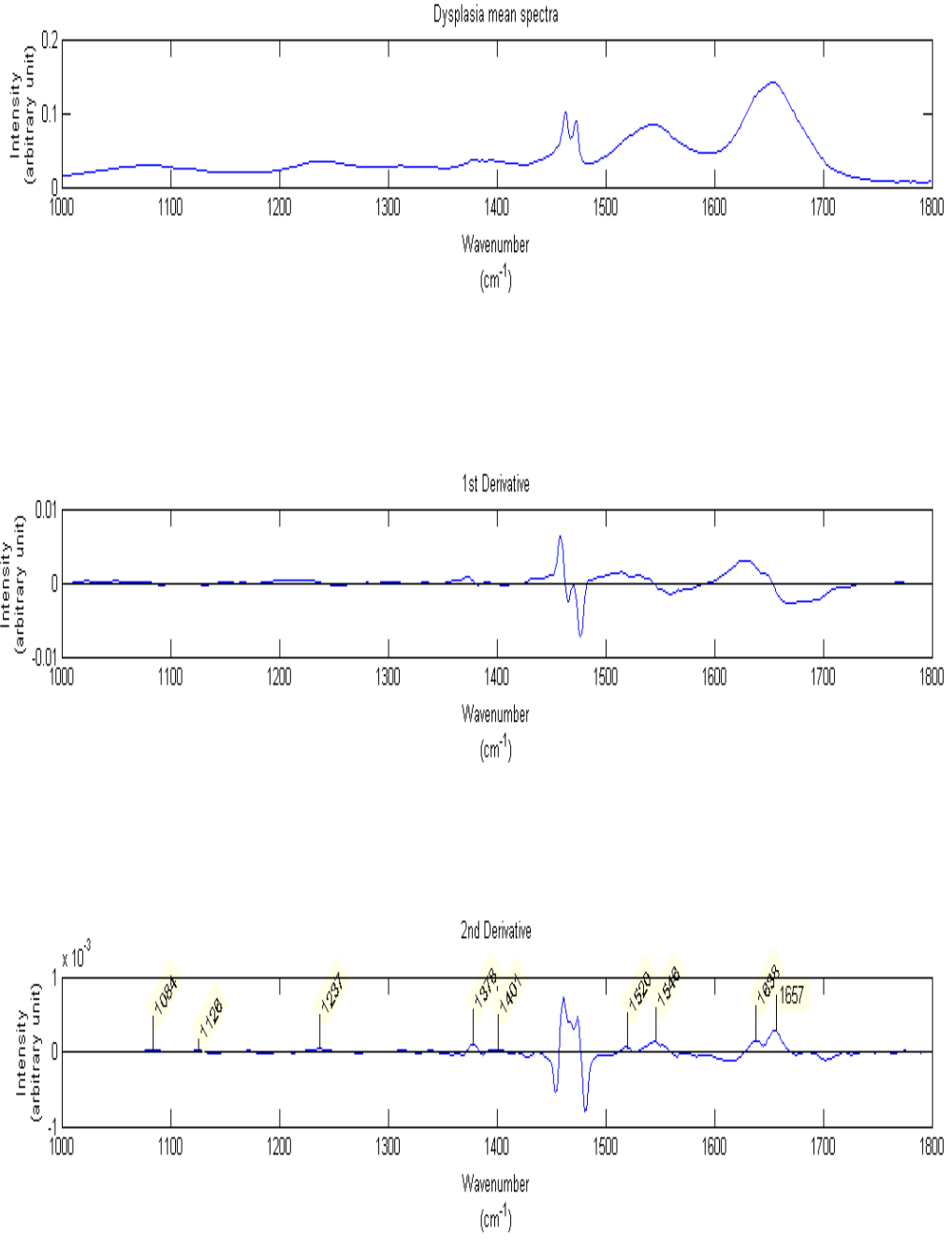
Figure 5.13: Dysplasia group derivatives; (Top) Mean spectra, (Middle) 1ˢᵗ order derivative, and (Bottom) 2ⁿᵈ order derivative for Dysplasia group
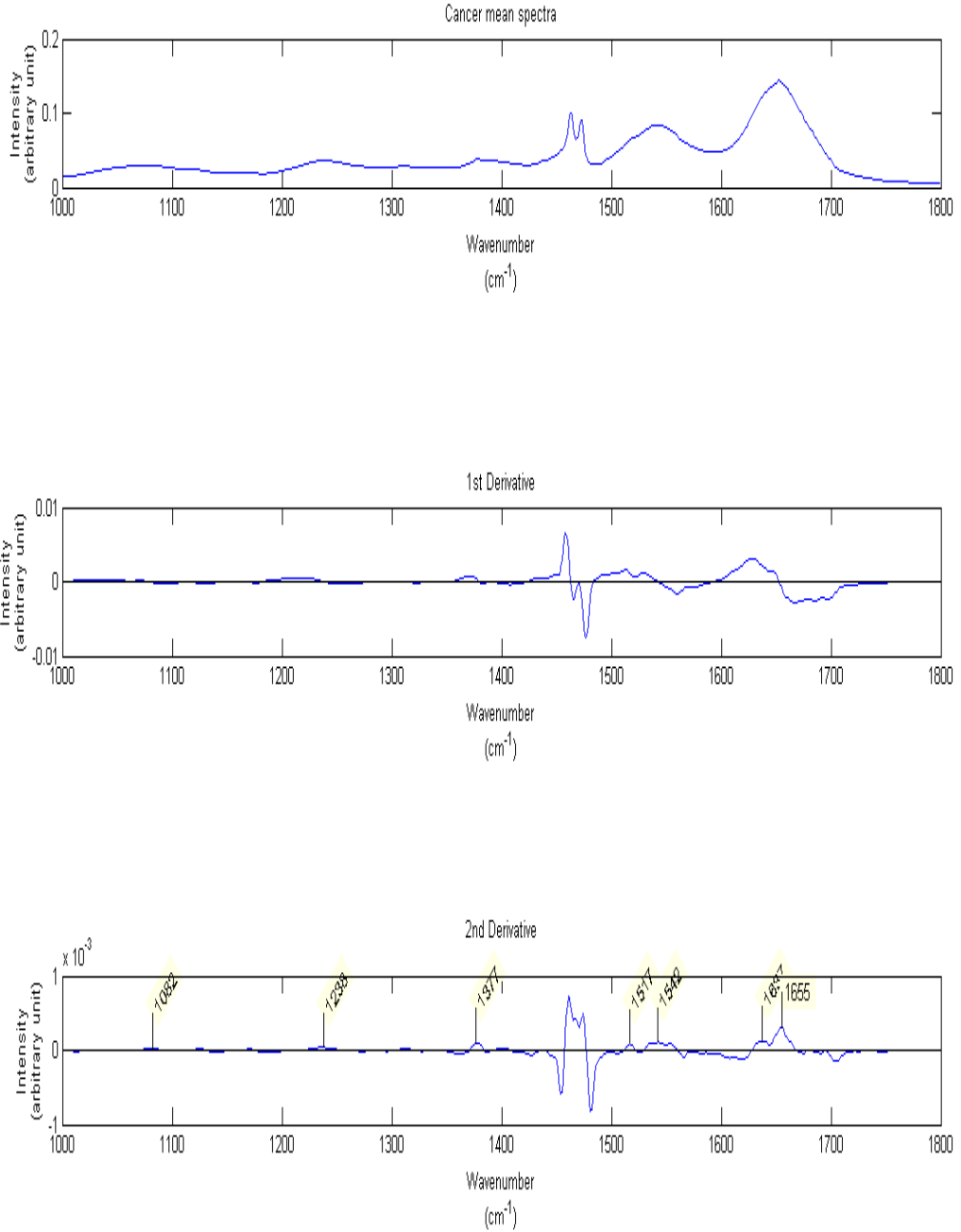
Figure 5.14: Cancer group derivatives; (Top) Mean spectra, (Middle) 1st order derivative, and (Bottom) 2nd order derivative for Cancer group
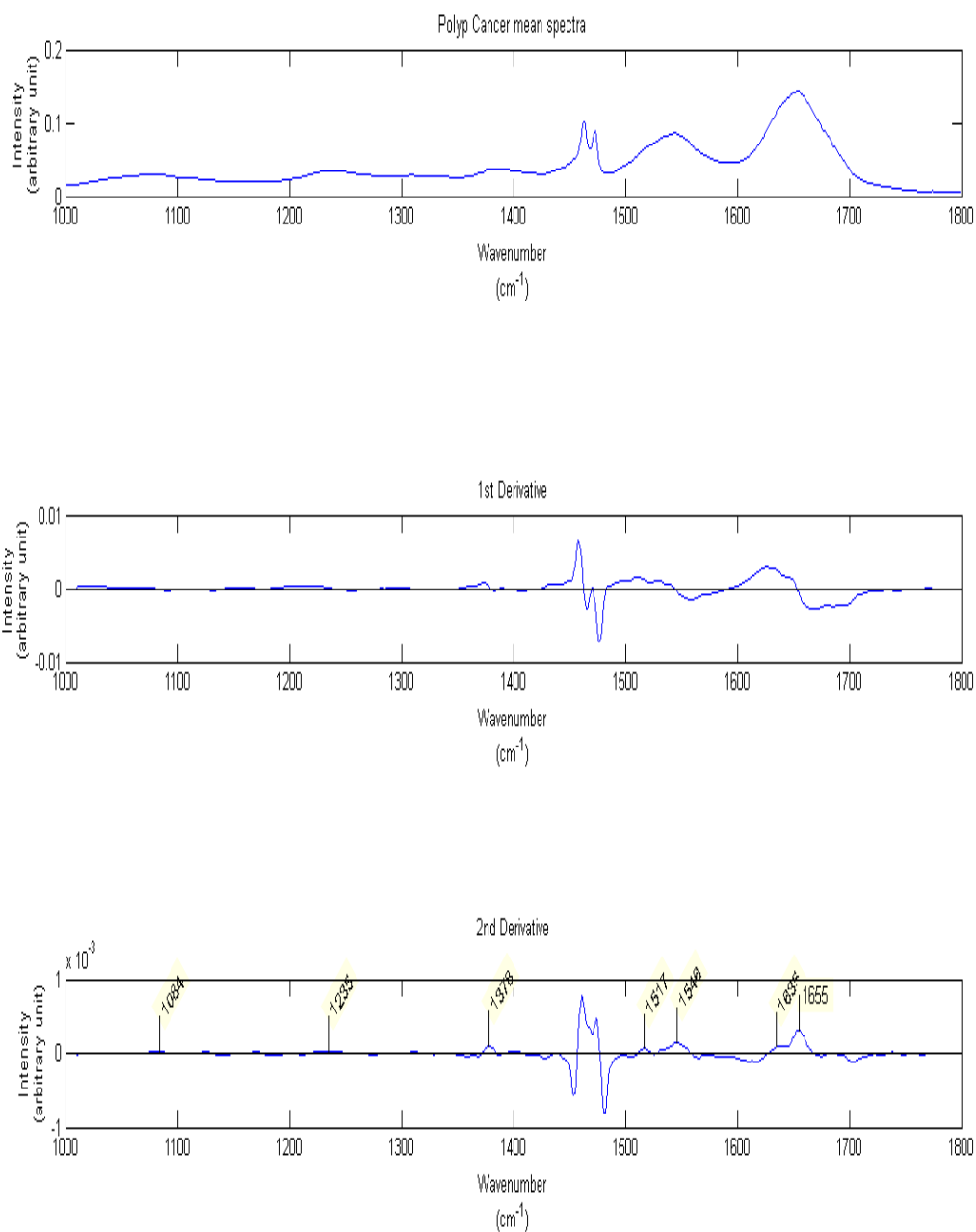
Figure 5.15: Polyp Cancer group derivatives; (Top) Mean spectra, (Middle) 1st order derivative, and (Bottom) 2nd order derivative for Polyp Cancer group
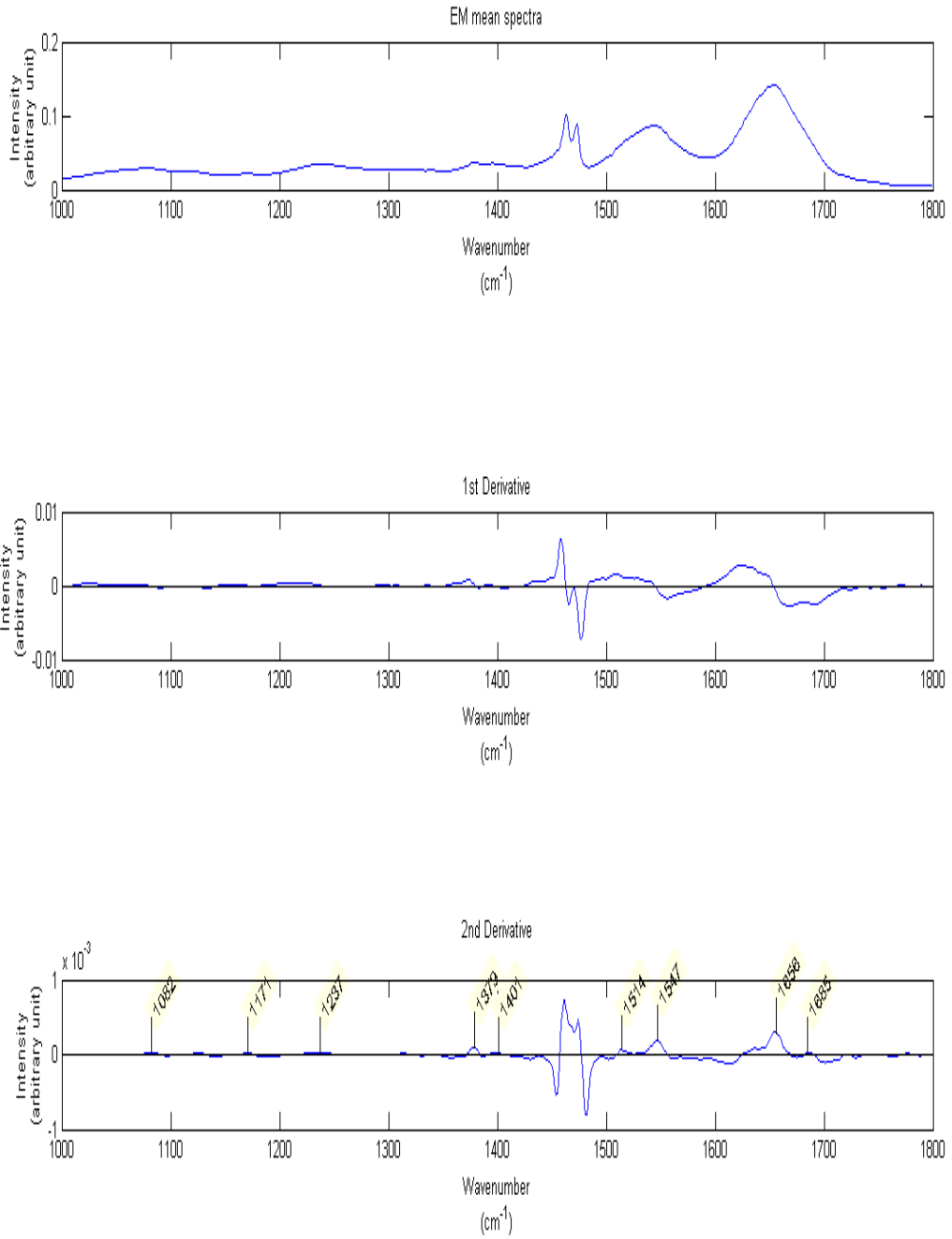
Figure 5.16: Epithelial Misplacement (EM) group derivatives; (Top) Mean spectra, (Middle) 1ˢᵗ order derivative, and (Bottom) 2ⁿᵈ order derivative for EM group

PO_2^- modes of nucleic acids and carbohydrates at 1080 cm^{-1} contribute to the infrared spectra and are observed in greatest intensity in normal tissue, although there is not much difference seen across the entire dataset, it characterises the C-O bending mode of carbohydrate (see Table 5.4). One group has suggested that this peak is also due to the heterogeneity of mucous present, and in particular of crypt mucus.(Krafft et al. 2008) It has also been attributed to DNA. Phosphate bands of phospholipid have also been found between 1080- 1230 cm^{-1}.

The peak at 1128 cm^{-1} is of greater intensity in normal and dysplastic tissue in comparison with the other groups. It has been assigned to mucin at 1121 cm^{-1}, with reduced concentration in diseased tissue because carcinogenesis cuts down mucus formation,(Conti et al. 2008) which would be in keeping with a much stronger peak observed in the normal and dysplastic groups. A lack of intensity in the EM group, as could be expected, could be because the internalised epithelium is not producing mucin to 'normal' levels seen on surface epithelium. This peak has also been attributed to RNA in normal tissue.(Argov et al. 2002)

A peak at 1171 cm^{-1} is observed in the EM group and not so in the other four. This is likely due to C-O stretching(Rigas et al. 1990) and may be due to a change in protein conformation secondary to a reaction to its submucosal environment.

The asymmetric PO_2^- stretching of nucleic acid exists at 1240 cm^{-1} can be attributed to collagen in connective tissue,(Conti et al. 2008) specifically of the submucosa,(Krafft et al. 2008) or to the phosphate group of DNA.(Krafft et al. 2008) The presence of this peak across all groups could indicate that as well as identification of epithelium, some submucosa was also involved, which would be particularly relevant in the invasive Cancer and EM groups given their glandular placement in the submucosa. This has also been attributed to an Amide III band in this position from C-N stretching and N-H bending. (Gasparri and Muzio 2003)

Symmetrical CH$_2$ deformation modes at 1400 cm^{-1} and asymmetrical CH$_3$ at 1460 cm^{-1} were presumed to be due to hypomethylation in cancer molecular structural change during carcinogenesis(Conti et al. 2008) or lipid contribution. It

Page | 98

is observed in the Dysplasia and EM groups (1401 cm^{-1} nearest reference peak), in keeping with Argov et al. who demonstrated various higher peaks in the Dysplasia groups and found that it correlated with grade of dysplasia.(Argov et al. 2002)

The most distinctive peaks in the IR absorption spectra are assigned to the absorption of proteins: and this can be seen even in the raw spectral data. Amide I and II peaks found across all groups from 1513 cm^{-1} to 1685 cm^{-1} (Li et al. 2005; Krafft et al. 2008; Wood et al. 2011) vary in nature; the whole region demonstrates change in breadth of peaks and shoulder region varies.

The most intense is the peak of Amide I found at 1655 cm^{-1} which corresponds to the C=O stretching with N-H bending and C-H of peptide bonds.(Ami et al. 2003) This derives from a α-helix conformation, which could arise from a change in the secondary structure of proteins during carcinogenesis.(Conti et al. 2008) The specific peak of Amide I at 1637 cm^{-1} is attributed to epithelial peptide groups, and potentially polyglycosylated peptides in normal mucus.(Krafft et al. 2008) It has been reported that the shape of this Amide I band is influenced by the secondary structure of proteins; α helices, β pleated sheets and other formations.(Lui et al. 2001)

A prominent Amide II band at 1546 cm^{-1} arises from C-H-N and C-N stretching of peptide bonds in different vibrational modes.(Gasparri and Muzio 2003)

5.2.2 Difference Spectra

Difference spectra allow the best comparison to be made between groups, where otherwise only very subtle differences in peak height and intensity might be present. This method of analysis allows emphasise to be placed on spectral variation and to extract differences from one group to the next. The initial peak assignments for raw mean spectra in all five pathological groups were the same across the groups except for 1650 cm^{-1} peak which demonstrated subtle change in peak height (wavenumber shift by only one or two).

Difference spectra highlights the net difference between the sample and reference state spectrum, without distortion of peak heights, widths or noise levels, which simplifies data interpretation. The method does not however

Page | 99

overcome the difficulty of assigning relative importance to changes in spectral bands. Rather, the magnitude of the changes in the sample spectrum relative to the reference spectrum is used to drive data evaluation and interpretation.(Vrettos et al. 2006)

The difference spectrum were calculated using an in-house MATLAB programme by subtracting mean normalized spectrum from one group from the same for another group. Figures 5.17- 5.20 show the resulting difference spectra obtained from taking the difference between mean normalised spectra from each group in turn against the Cancer reference group. The positive peaks in each plot represent biochemical constituents in abundance in the Cancer group, in contrast the negative peaks which represent lower concentration in the opposing Group. The noisy regions are removed from interpretation, and only the wide peaks are labelled and considered for description of the net group differences, with analysis based on the description of their size and position.
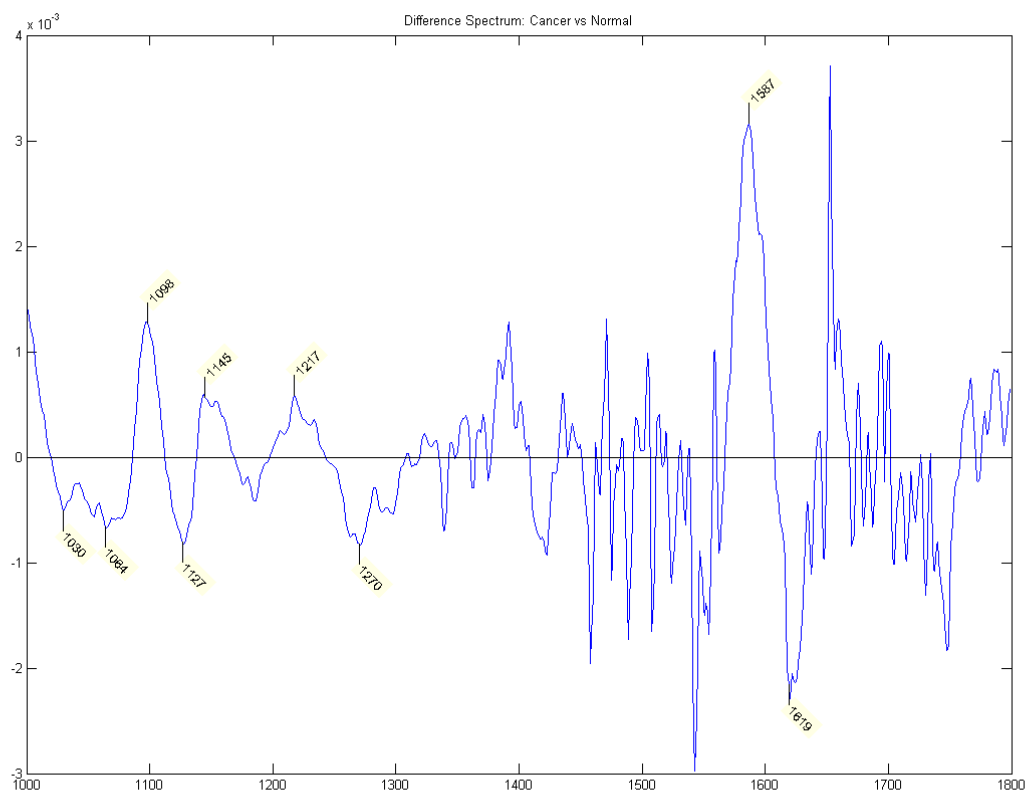
Figure 5.17: Cancer vs. Normal difference spectra

The comparison between Cancer and Normal groups in Figure 5.17 demonstrates the strongest peak at 1587 cm^{-1} most significant in the positive region, so in the Cancer group. This is in keeping with the previous derivative findings that the protein Amide I and II change in intensity, probably due to change in structure, in the carcinogenesis process. That peak is situated in a wide band of noise from 1300-1800 cm^{-1} which doesn't provide other useful information for those wavenumbers because the peaks are so sharp. This will have some paraffin contribution also complicating that section. Peaks at 1098 cm^{-1} and 1217 cm^{-1} attributed to nucleic acids and fatty acids are greater in cancer compared with normal whereas short broad DNA peaks at 1030, 1060 cm^{-1} along with RNA at 1127 cm^{-1} are stronger in the normal group.
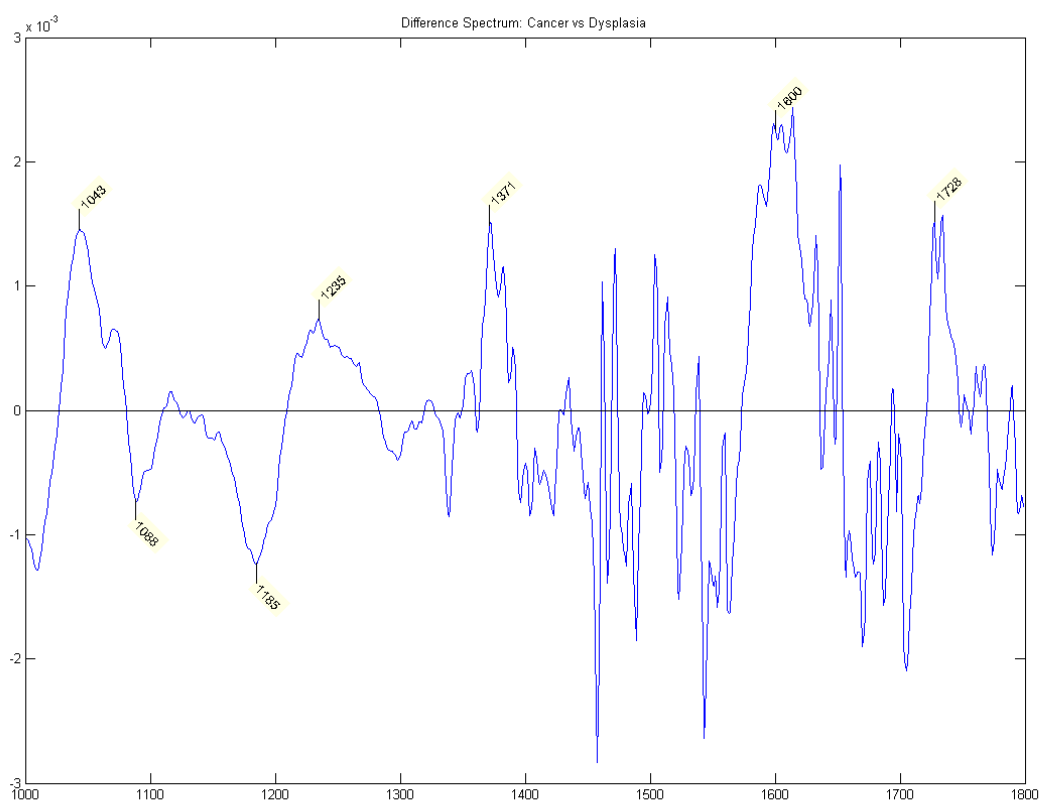
Figure 5.18: Cancer vs. Dysplasia difference spectra
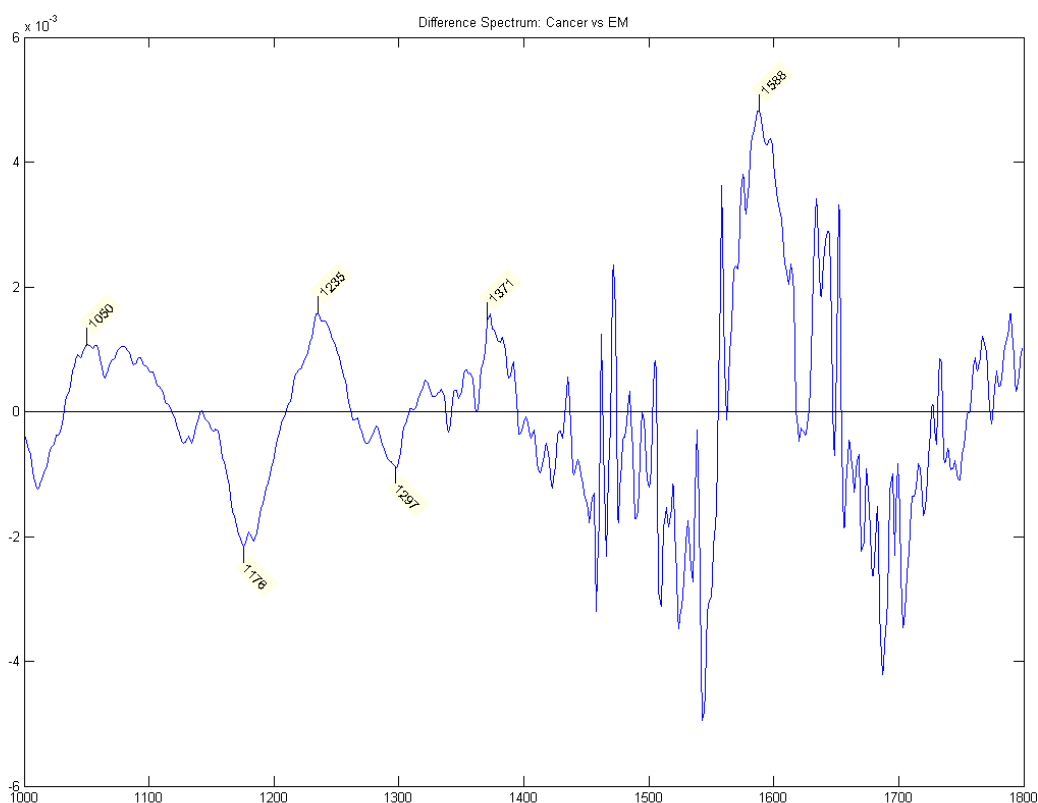
Figure 5.19: Cancer vs. Epithelial Misplacement difference spectra

Figures 5.18-5.19 also demonstrate noise in the region of 1300-1800 cm^{-1} with very limited useful peaks, however, it appears that an intense, broad peak at approximately 1600 cm^{-1} appears to be most significant in the Cancer group compared to every other group, equally fatty acids at 1235 cm^{-1} and DNA at 1371 cm^{-1}. Although smaller peaks are most significant in the cancer group compared with all others. Glycogen at 1040 cm^{-1} is significant in cancer compared to all other groups except for normal, where the peak is not prominent, which would be in keeping with a modified glycogen and energy storage facility in carcinogenesis (the cancer process should be using energy stores).
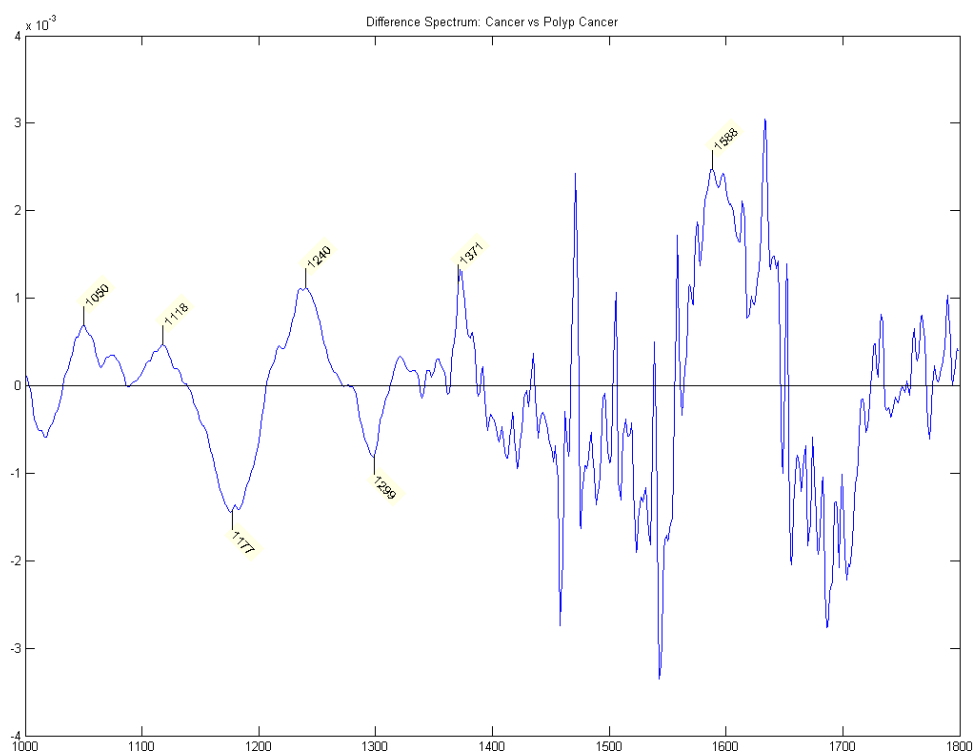
Figure 5.20: Cancer vs. Polyp Cancer difference spectra

Peak differences between Cancer and Polyp Cancer groups show the least differences in overall intensity demonstrated. Although there are differences present between the two, they might indeed be very small and subtle, as would be expected. It could also be argued that the 1588 cm^{-1} peak is also very noisy, and therefore not reliably useful as a peak of significance to differentiate between the two groups.

5.3 Bivariate Analysis

Peak intensity ratio calculations are performed to better evaluate the biomolecular changes present between different groups. They reflect the ratios of protein, nucleic acid (NA) and fat content of the cells present and therefore give an indication of the carcinogenesis process, see Table 5.4. This has been used by many groups to demonstrate their findings.(Rigas et al. 1990; Li et al. 2005; Zwielly et al. 2010) In this case the sum of the peak area of interest was

used (instead of the intensity at the exact wavenumber) using inverse 2nd derivative mean spectra.

Group Peak Area Ratio	NA: Protein (1038-1112:1610-1692)	NA: Lipid (1038-1112:1375-1383)
Normal	**0.069**	0.318
Cancer	0.106	0.500
Dysplasia	0.09	0.467
Polyp Ca	0.093	0.484
EM	**0.065**	0.366

Table 5.4: Sum of peak area ratios for Group comparisons

*NA= nucleic acids

It is to be expected that during the process of carcinogenesis the cell lipid content is low because the energy is being used up for protein synthesis. Nucleic acids are large polyatomic molecules which include DNA and RNA made from nucleotides. Each nucleotide is made from a 5-carbon sugar, a phosphate group and a nitrogenous base. The sugar can be either oxyribose or deoxyribose according to its destination for RNA/ DNA. Therefore nucleic acid content should be associated with high protein production and lipid consumption with active carcinogenesis.(Isabelle et al. 2008; Wood et al. 2011)

The NA:Lipid peak ratio is highest in Cancer, Polyp Cancer and Dysplasia groups. This corresponds with highest ratio scores for NA:Protein content in Cancer and Polyp Cancer relative to Normal and EM, which also correlates with discriminatory peaks described earlier. It should be expected that Normal and EM would be of similar value, the fact that cancers are found to be different

would assist in the clinical scenario discriminating either from cancer. The greatest discrimination appears to be provided by the NA:Protein ratio which may therefore be useful to help separate the EM Group from the other groups. Interestingly the Dysplasia group is mid-table for both ratio comparisons which would be in keeping with the progressive change in cellular biochemistry from normal to malignant and associated biochemical changes described, although Argov et al. found highest peak ratios for NA and lipid overall in their dysplastic groups.(Argov et al. 2002)

The ratios used here are calculated from the mean spectra with overlapping standard deviations therefore further work would be needed to be able to produce individual error bars for these values and confirm the results found.

5.4 Multivariate Spectral Analysis

Univariate and bivariate analysis techniques are useful in analysis of individual peaks and to analyse the differences between groups in descriptive terms, however as previously discussed, there is considerable overlapping of peaks and band assignments which makes the techniques less favourable than using multivariate analysis to really tease out the differences and assess the importance given to the differences between the groups. For these reasons it was important to implement a combination of both supervised and unsupervised multivariate techniques to construct a classification model based on variance observed between the spectra. The use of Principal Component Analysis (PCA), an unsupervised method, then Linear Discriminate Analysis (LDA), a supervised method, is well described in the literature. Leave One Sample Out Cross Validation (LOSOCV) was then subsequently used to probe the performance of the classification model. The LOSOCV is employed as a robust test for the model. From this, sensitivity, specificity and overall performance could be determined.

In this analysis a 'Four Group' and a 'Two Group' model were performed and these will be discussed at length in the following sections. At this point a note should be made about the Polyp Cancer and Cancer groups. It was found that the Difference and first and second Derivative spectra for these two groups demonstrated a high degree of overlap (see Figure 5.15 and 5.19), and it was

Page | 105

difficult to separate them out on the basis of the early analysis of the spectral biochemical information. This mirrors our biochemical understanding that there should be a high degree of overlap between early polyp cancer and advanced cancer, because there is little biochemical difference between the two except for that of histological Grade (variable low-high). Any differences demonstrated between the two may be due to 'invasiveness' since that is the only discernible difference between the two, an unidentified molecule such as a protein may be found in higher quantity in the Cancer group because it is invading through the submucosa to a varying degree away from its position on the epithelial surface, whereas Polyp Cancer group may be lacking in this because it has not demonstrated any invasion through the submucosa. To assess whether there was truly only a very small difference between the two, PCA-fed LDA analysis was performed of the Polyp Cancer and Cancer groups in comparison to the Normal group, to assess the expectation that it would be difficult to separate the two cancer groups against the normal spectra and validate the combination of the two groups.

PCA-LDA for the cancers against normal spectra demonstrated that the specificity of the model to separate normal from cancer is higher (88%) than the specificity to discriminate Cancer from polyp cancer (76%). These results justify the combination of the two cancer groups together for analysis as one (therefore increasing the group size by default and increasing the spectral information for the model).

5.4.1 'Four Group' Model

The Four Group Model dataset consists of individual spectra from Normal, Dysplasia, and Epithelial Misplacement groups and the Polyp Cancer and Cancer groups together as Combined Cancer, labelled 'Abnormal group' (a total of four groups all together) A comparison of the mean spectra between the four groups can be seen in Figure 5.21 with standard deviation plotted. The mean spectral data demonstrates several differences in the peaks between the four group means, particularly in the regions of 1080 and 1240, and 1550 and 1650cm^{-1}, as discussed previously. However there is overlap in the standard deviation of the group means, particularly noticeable at 1550 and 1650cm^{-1}and

the trough of 1600cm[-1], so there is some significant spread of the values at these points. Because the SD is based in the mean, the spread cannot be attributed to the whole group or one aberrant sample within the group.
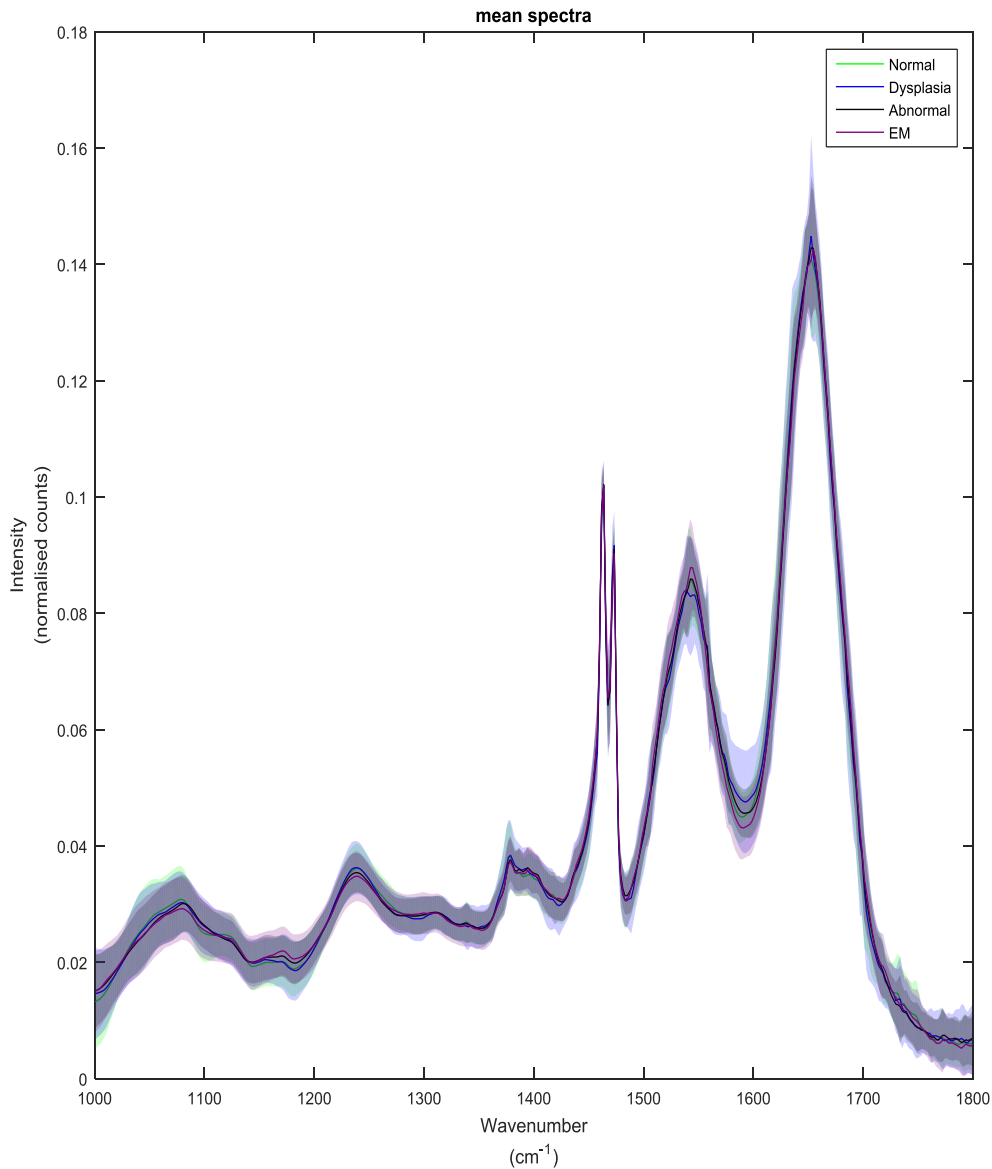
Figure 5.21: Graph demonstrating mean spectral differences in Four Group Model. Green = Normal group, Blue = Dysplasia group, Purple = Epithelial Misplacement group and Black = Abnormal combined cancer group with standard deviation plotted as shaded areas

To establish biochemical changes in the spectral dataset across the selected epithelial spectra selected, Principal Component Analysis (PCA) was used. PCA reduces the whole dataset to a few spectral components and reveals the maximum variance across the whole dataset. The PCA creates PC scores and loads in order of variance. Analysis of variance (ANOVA) was then performed on the PCs to identify the most significant PCs, a histogram of which can be seen in Figure 5.22 demonstrating the importance of each individual PC (of the top 25). In the Four Group Model, PCs 2, 3 and 4 (with significant contribution from PC1 and 6 too) were the top three most significant PCs. Throughout the experimental process there are many factors which could erroneously increase variation between the groups such as;

- Paraffin signal despite good EMSC correction within the programme
- Storage of slides- scratches from transport and dust and dirt from cardboard storage sleeves
- Experimentation set-up variation
- Instrument variation (light source)
- Cell cycle variation within a pathological group
- Environmental factors such as water and carbon dioxide vapour
- Misclassification of glands during identification process

These are all acknowledged and as much as possible corrected at the time so that the PCA scores can be relied upon to be a true measure of variation in tissue rather than anything else. However corruption by extraneous factors could still contribute to the separation, hence the need for a detailed review of each PC loading which thus aims to avoid the potential for any of them to be significant in calculation the variation across the dataset.
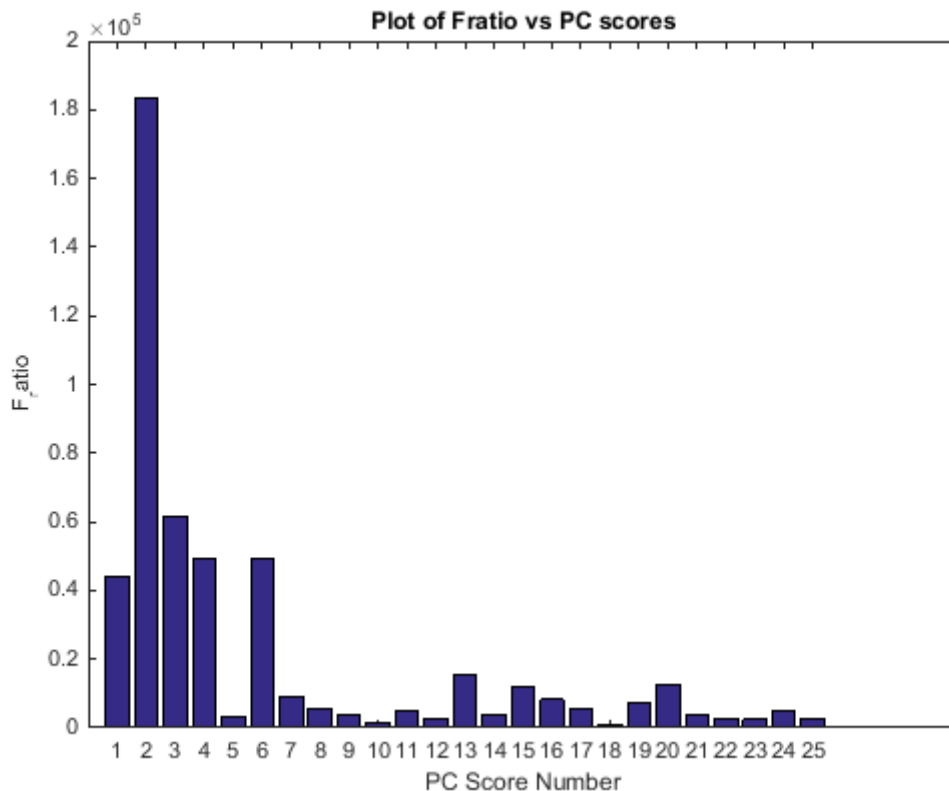
Figure 5.22: Histogram to show plot of Fratio vs. 25 Principal Components score

A scatter plot was also created to demonstrate the maximal variance of the groups in three dimensions (3D); using the three most important PC selected plots (in this case 2, 3 and 4). It generates an image of how well the groups separate on the basis of three PCs in the context of 3D time and space separation, see Figure 5.23. The separation of the Four Group Model appeared good, with some overlap as expected but the majority of the groups being well circumscribed and distinct from each other. The Dysplasia group demonstrates the greatest inter-group variability with wide spread and the most outliers, probably because of the significant pathological variation within the group, ranging from near-normal to near cancer and all in between. This is to be expected considering that some glands will close to normal biochemically and some much more similar to a Cancer type.
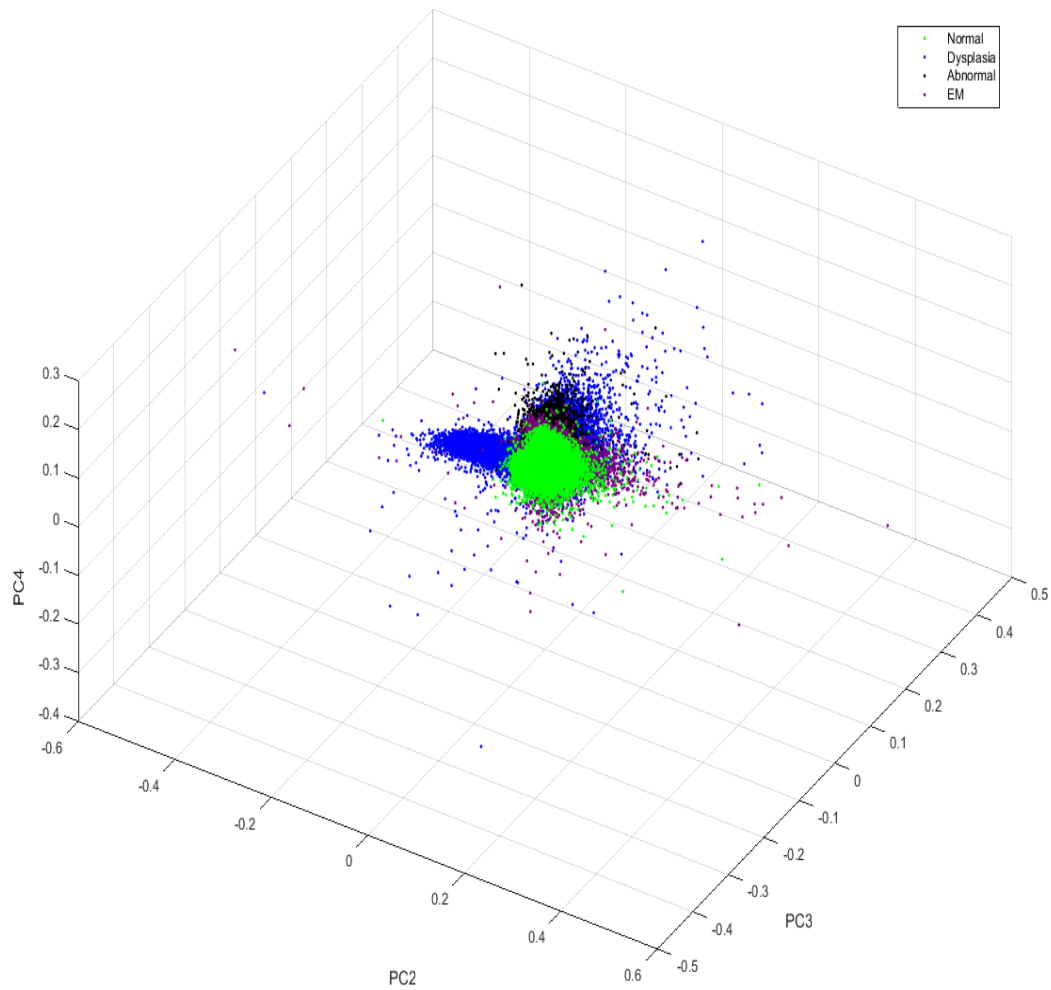
Figure 5.23: PCA Scatter Plot (PC2 vs. PC3 vs. PC4) demonstrating maximal variance in Four Group Model. Green = Normal group, Blue = Dysplasia group, Purple = Epithelial Misplacement group and Black = Abnormal combined cancer group

Figure 5.24 demonstrates the PC loadings for the most significant of the three principal components PC 2, 3 and 4 in detail:
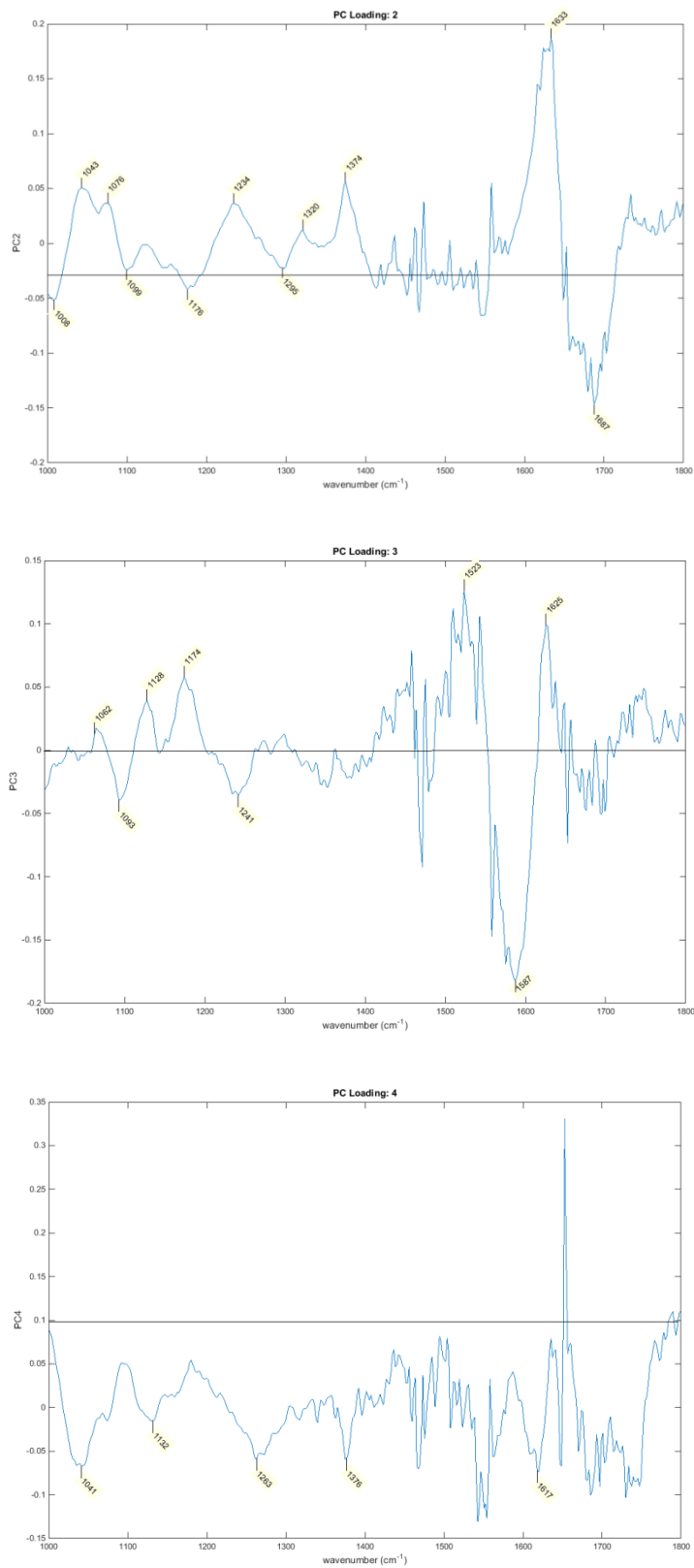
Figure 5.24: PC loadings for PC 2, 3 and 4 with peak label assignments

Peak assignment labels have been added to these PC loadings (Figure 5.24). The most significant peaks in the 3 PCs are a recurrence of 1040, 1095, 1170, 1240 and 1630cm^{-1} as discussed previously.

These PCs were then used to develop the Linear Discriminate Analysis (LDA) models for the Four Group analysis. All 25 of the calculated PCs were used as input into an LDA classification to maximise the variance between pathological groups and reduce variance within groups. A two dimensional scatter plot of the linear discriminant (LD) scores for the dataset are displayed in Figures 5.25. Each dataset produced various combinations of LD subplots as a result of the LDA performed which have not been included in this thesis due to sheer volume of data, but have been examined as part of the analysis.

The efficiency of the PCA-fed LDA classification model was then evaluated using leave one sample out cross validation (LOSOCV) described in Chapter 4. This method determines whether the predications of the classification model are accurate, is independent data (i.e. new case samples) is not available. The classification model was constructed with all 25 PCs and results reflect the number of spectra correctly classified for each group within the Four Group Model, see Table 5.5.

Pathology Group	Spectral Classification			
	Normal	Dysplasia	Combined Cancer	EM
Normal	**85033**	39384	36712	17037
Dysplasia	93094	**199827**	161559	25115
Combined Cancer	73969	110987	**356515**	106355
EM	69107	29865	62893	**238575**
Average Sensitivity (%)	48	41	55	60
Average specificity (%)	85	85	75	89

Table 5.5: Leave one sample out cross validation confusion matrix table with average sensitivity and specificity. EM= Epithelial Misplacement.
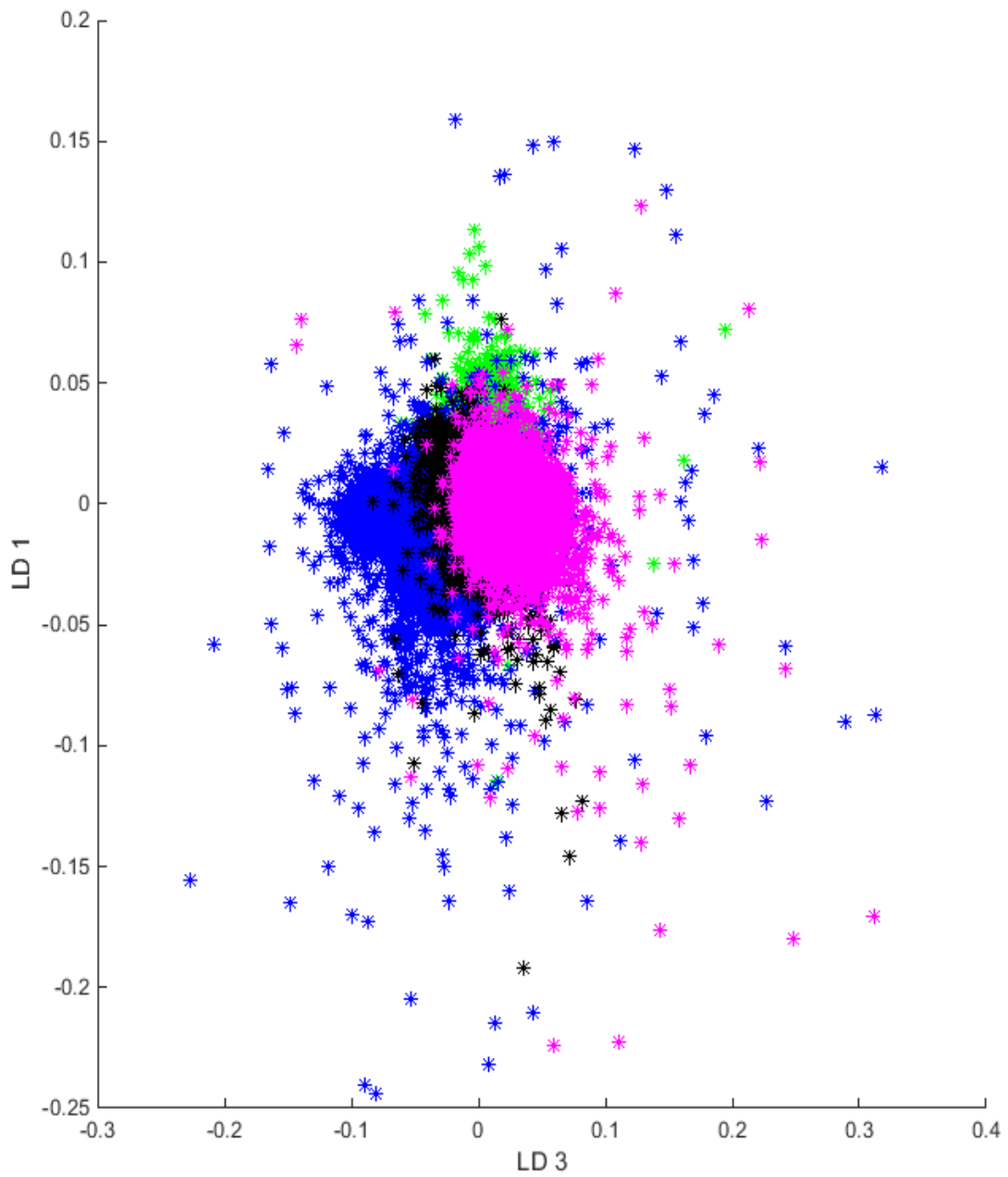
Figure 5.25: LDA Scatter Plot (LD1 vs. LD3) represents the clustering of pathological groups for the Four Group Model; each point represents a single spectra. Green= Normal group, Blue= Dysplasia group, Black= Abnormal group and Purple= Epithelial Misplacement group

These results show a high average specificity (75- 89%) but low average sensitivity (41- 60%) in comparison for the LOSOCV evaluation. This reflects a poor measure of PCA-fed LDA classification model performance. This large number of misclassified spectra may be due to unbalanced number of spectra in each group, significant differences in cell sample with variation in degree of maturation and cell cycle, problems with epithelial cell selection, inclusion of poor quality spectra, and inter-sample spectral variability within the groups. Ideally the number of spectra per group in the model should be equal for the LOSOCV procedure to avoid poor prediction results. It was decided to perform the whole process again with the same number of spectra in each group to see if this had a significant effect on results. 178,000 spectra (smallest group determines number which was the Normal group) were selected at random from each group and entered into the PCA-LDA model for LOSOCV. The results of which can be seen in Table 5.6:

	Spectral Classification			
Pathology Group	Normal	Dysplasia	Combined Cancer	EM
Average sensitivity (%)	55	41	55	61
Average specificity (%)	85	86	78	89

Table 5.6: Leave one sample cross validation sensitivity and specificity results for the revised Four Group Model using only 178000 spectra per group. EM= Epithelial Misplacement

This normalisation for spectral number made a small improvement to classification of the Normal group but overall minimal significant improvement, therefore the poor classification model is probably explained by pathological variation and selection error, rather than group size.

It was decided to further analyse the model performance by using the percentage of classified spectra in each specimen, rather than just the spectra alone out-of-context with the pathological classification. This was performed by classifying each case per group depending on majority spectral classification to give a 'majority spectral vote' per case. This was to better reflect the

Page | 114

classification per sample and remove the bias of outlying spectra which may have affected LOSOCV performance initially. The classification table is shown in Table 5.7.

	Case classification by majority percentage			
Pathology Group	Normal	Dysplasia	Combined Cancer	EM
Normal	**11**	2	2	0
Dysplasia	2	**9**	1	1
Combined Cancer	2	3	**10**	4
EM	1	2	3	**8**
Average Sensitivity (%)	73	69	53	57
Average specificity (%)	89	85	86	89

Table 5.7: A confusion matrix table based on classification of cases per group by majority percentage classified spectra, with average sensitivity and specificity. Correctly classified cases in bold. EM= Epithelial misplacement

Table 5.7 demonstrates a much improved model performance when analysed on a percentage basis rather than for all spectra individually. Particularly sensitivity is much improved across the pathology groups, and prediction greatly improved (reading across the rows). In the normal group, only 2 were misclassified into dysplasia and combined cancer respectively, 4 samples out of the dysplasia group spread across the other groups. The cancers and EM cases were most frequently misclassified into each other's group, which might be expected given that histologically these are the most difficult to distinguish from each other, and the differences therefore are very small for the model to use. The majority vote acknowledges that there may be significant number of misclassifications per sample but, if overall, the classified spectra put the sample in its true group then the model has performed well, hence the increase in performance. This probably reflects a true comparison with a pathologist discriminating between different pathology cases where there may be evidence of mixed glands present or uncertainty in difficult cases.

5.4.2 'Two Group' Model

The next step in the analysis of the spectra was to assess the use of multivariate analysis in the differentiation of EM and Cancer groups. As discussed in Chapter 1, this is an area of great clinical need (see Section 1.7.1) with great scope for the use of infrared analysis to differentiate cancer from non-cancer as an additional complimentary tool to aid accurate and efficient diagnosis.

The Two Group Model dataset consists of individual spectra from Combined Cancer (Polyp Cancer and Cancer groups together) labelled 'Abnormal' group and Epithelial Misplacement group. A comparison of the mean spectra between the two groups can be seen in Figure 5.26 with standard deviation (of the mean) plotted. It demonstrates several differences in the peaks between the two groups, particularly in the regions of 1080 and 1550 cm^{-1} as discussed previously, although overlap in the SD of the population means is also demonstrated making it difficult to assess their significance based on this alone.
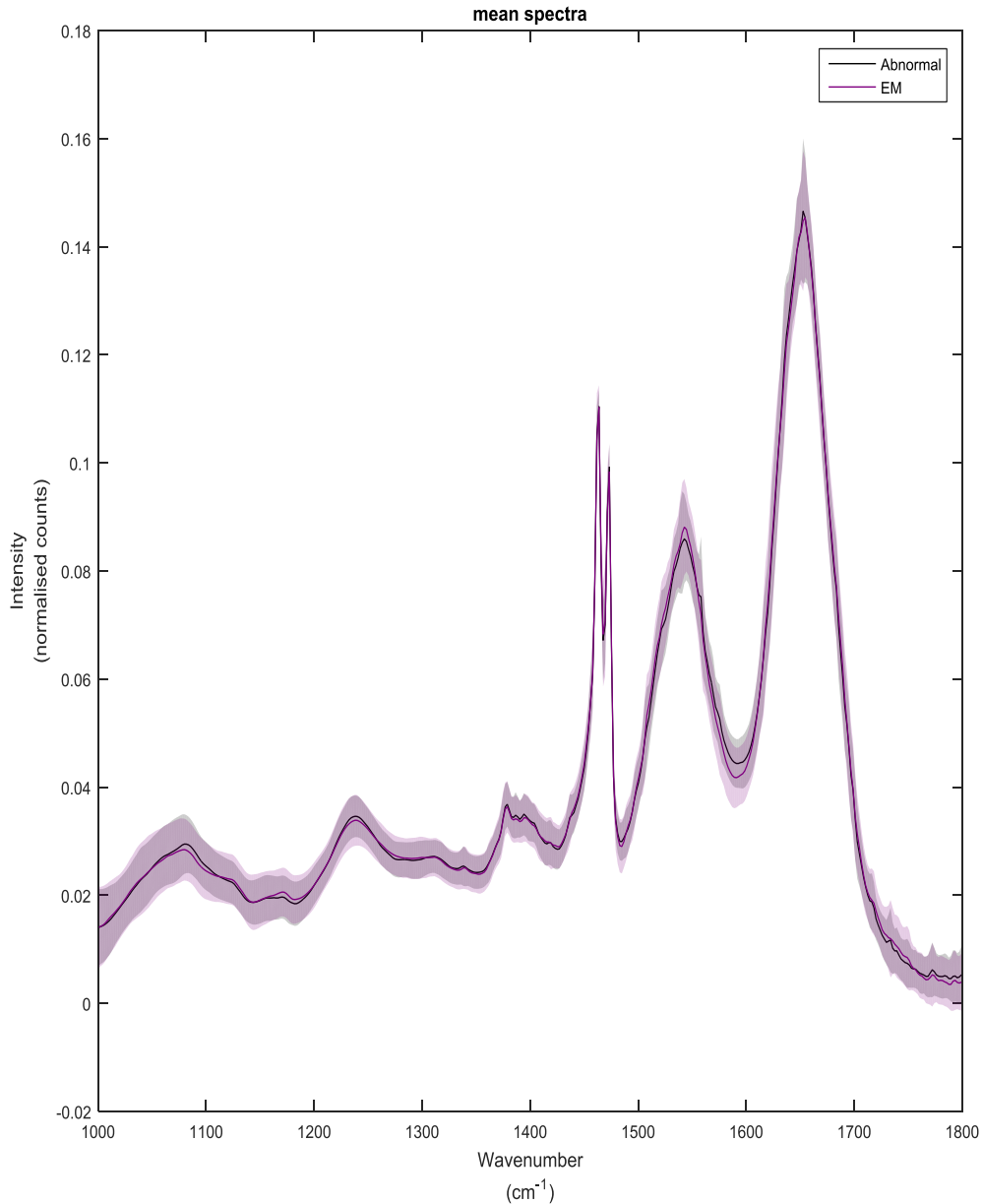
Figure 5.26: Graph demonstrating mean spectral differences in Two Group Model. Purple= Epithelial Misplacement group and Black= Abnormal group with standard deviation plotted as shaded areas

PCA was performed for the Two Group Model to determine the variance across the whole dataset. Again, the top 25 most significant PCs were determined, ordered and listed by ANOVA. Figure 5.27 demonstrates very clearly that there were three PCs, 2, 3 and 4, which were much more contributory than the others in the top 25 identified.
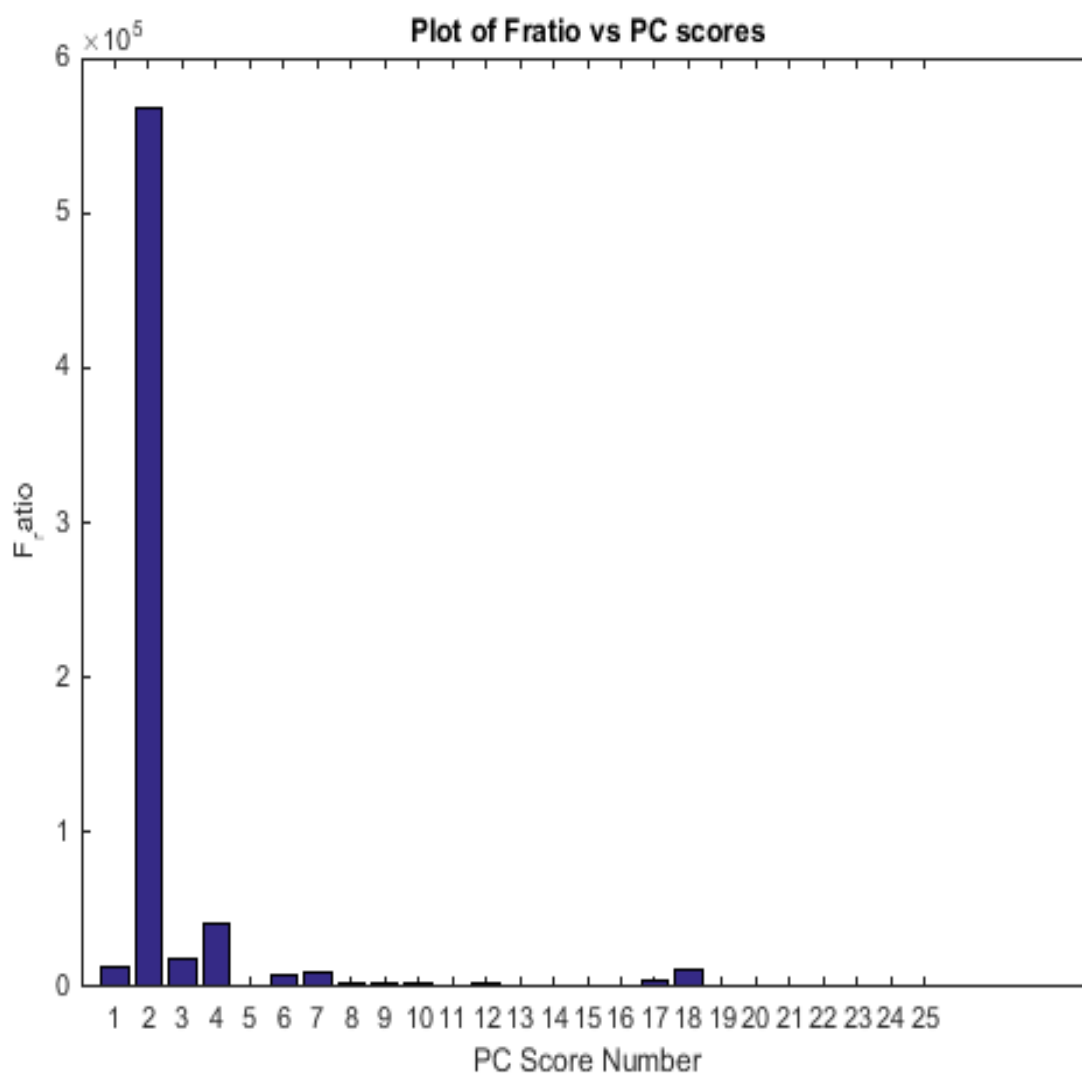
Figure 5.27: Histogram to show plot of Fratio vs. 25 Principal Component scores

The three dimensional scatter plot (Figure 5.28) using PC 2, 3 and 4 demonstrated a clear separation between the two groups, which although adjacent, shows two well described populations in separate clusters containing minimal outlying spectra, (given that every spectra is plotted), implying that the epithelial identification was strong in this selection of spectra with minimal inter-group variability. All PC loadings were examined but Figure 5.30 demonstrates the most significant PC loadings, PC 2, 3 and 4, in more detail, with identification of major peak assignments, particularly repeating the theme of 1080, 1230, 1580 and 1650 cm^{-1} as previously discussed.
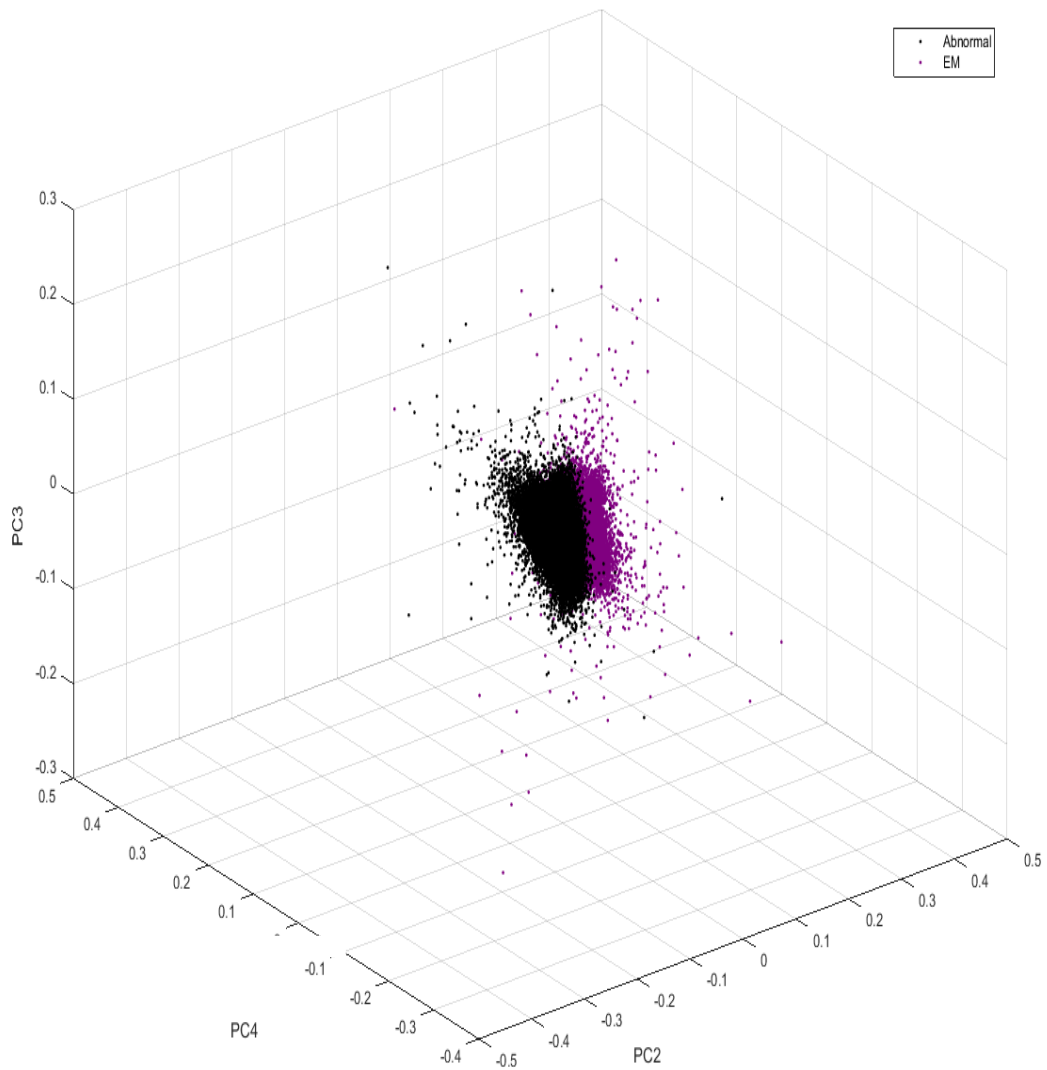
Figure 5.28: PCA Scatter Plot (PC2 vs. PC3 vs. PC4) demonstrating maximal variance in Two Group Model. Purple= Epithelial Misplacement group and Black= Abnormal combined cancer group
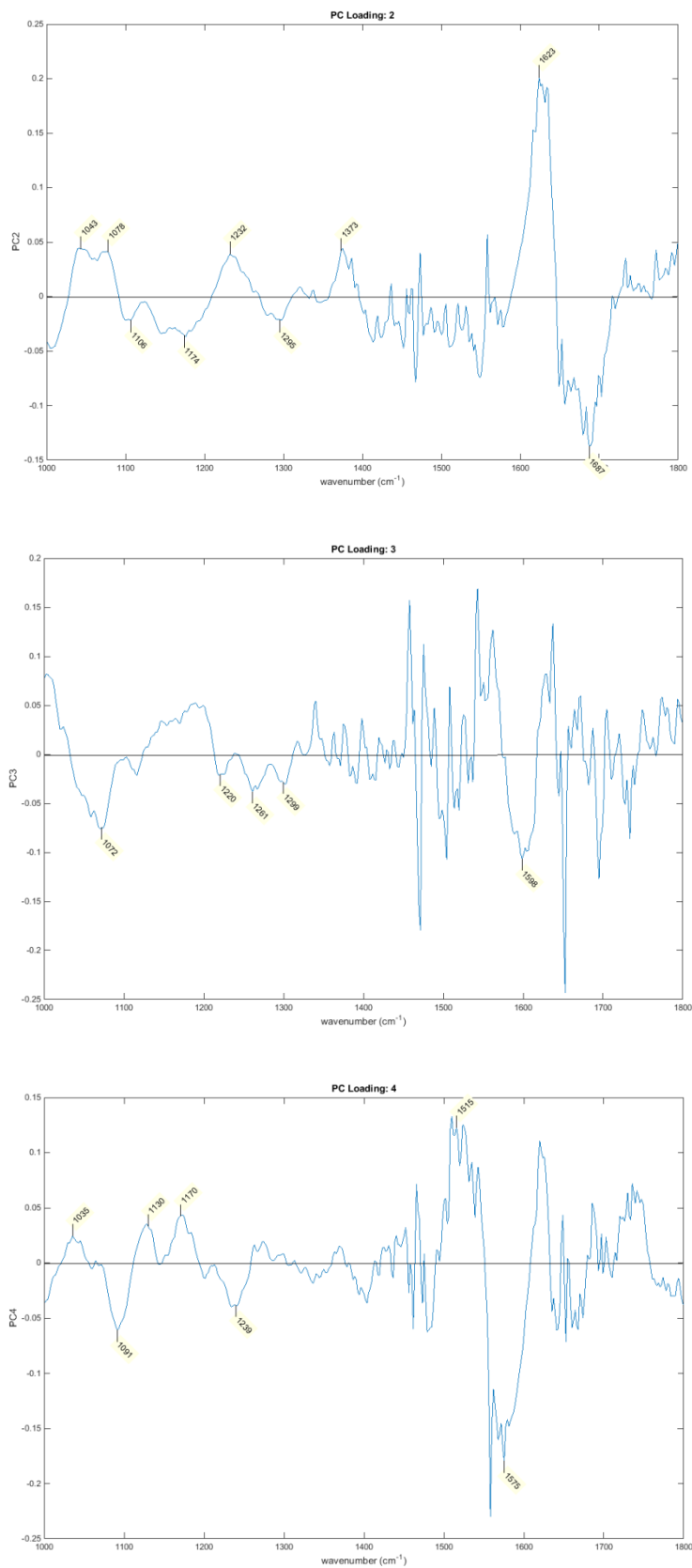
Figure 5.29: PC loadings for PC 2, 3 and 4 with peak assignment labels

The PCs were then used to develop the LDA model for the Two Group analysis, and a LDA histogram representative of the group clustering produced, see Figure 5.30. This shows excellent group separation, with only very minimal overlapping at the base. This very clearly shows that the PCA-fed LDA model has been able to discriminate between the two pathology groups very effectively.
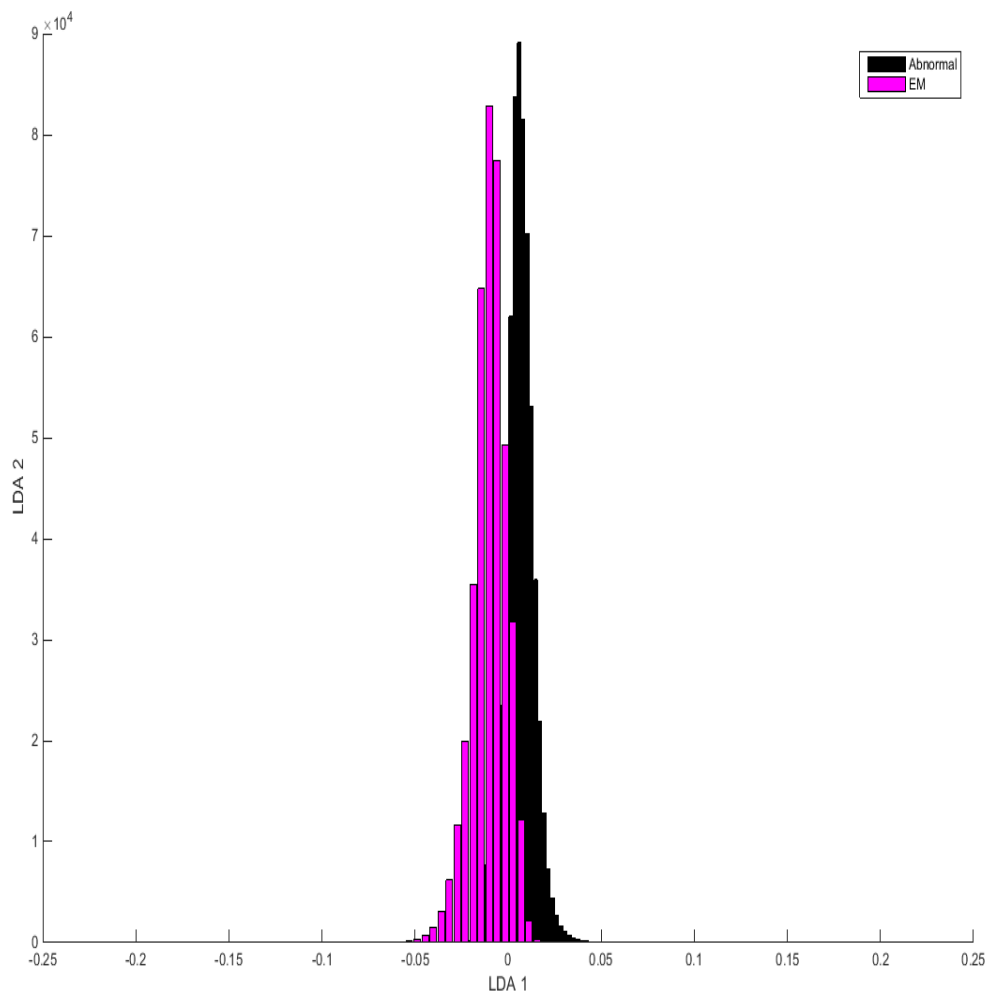
Figure 5.30: LDA Histogram (LD1 vs. LD2) representing the clustering of the two groups in the Two Group Model; Purple= Epithelial Misplacement group and Black= Abnormal group

The efficiency of the PCA-fed LDA model was then evaluated using LOSOCV again, reflecting the number of spectra classified correctly within the Two Group Model. The results of this can be seen in Table 5.8.

	Spectral Classification	
Pathology Group	Combined Cancer	EM
Combined Cancer	**533036**	114790
EM	101714	**298726**

Table 5.8: Leave one sample out cross validation confusion matrix.

EM= Epithelial Misplacement

The LOSOCV results are excellent for the Two Group Model, with an average sensitivity of 75% and specificity of 82%. This classification model was therefore highly successful at identifying the true positives with few misclassifications.

	Case Classification by percentage	
Pathology Group	Combined Cancer	EM
Combined Cancer	**16**	3
EM	3	**11**

Table 5.9: A confusion matrix table based on classification of cases per group by majority percentage classified spectra. Correctly classified cases in bold. EM= Epithelial misplacement

Again the PCA-fed LDA LOSOCV results were analysed by majority percentage on a per-case-basis per group as can be seen in Table 5.9. Again this has shown an overall improvement in the model performance with an increase in average sensitivity to 79% and specificity to 84%. This improvement is less dramatic than in the Four Group Model, because there were fewer misclassifications initially and the model was a higher performer. However it does demonstrate that analysing on the basis of percentage vote drives a

majority decision that is even more accurate for this Two Group classification model, an excellent result.

CHAPTER 6

CONCLUSIONS

The scope of this thesis was to evaluate the use of high resolution infrared spectroscopy in the analysis of colonic pathology. There is limited use of this novel technique in the literature, although much already at standard resolution for comparison. This research specifically investigated the role of high resolution infrared spectroscopic analysis in the differentiation of cancer from the benign pathological phenomenon of epithelial misplacement (EM), through the interrogation of the molecular structure of the various relevant pathological entities. There is conundrum in the diagnosis of epithelial misplacement and therefore huge potential for any novel analysis technique to have a positive impact in the diagnostic process in terms of speed and accuracy. Multivariate statistical methods, to extract the useful diagnostic information, were employed to discriminate normal colorectal epithelium from dysplastic and cancerous epithelium, as well as the discrimination of EM from cancer. Specifically principal component-fed linear discriminate analysis (PCA-fed LDA) was used to interrogate the experimental spectra and leave one sample out cross validation (LOSOCV) used to examine the robust quality of the classification.

The high resolution technique took approximately 55 minutes per sample to collect a huge amount of raw data; each tile contained 128x128 pixels (64 scans per pixel acquired), thus generated 1,048,576 scans per tile, a total of 9,437,184 scans (for 3x3 tiles) for analysis per sample. The samples were prepared using formalin fixed paraffin embedded tissue as per routine clinical preparation with no extra processing steps required. This allows the technique to be applied for use in clinical diagnosis, given that it does not require any extra processing and is a fairly rapid technique. This makes it supremely suitable as a complimentary diagnostic technique for difficult cases.

The high resolution spectra acquired from the Normal, Dysplasia, Cancer and EM pathologies demonstrated various differences in their biochemical constituents. Univariate and bivariate analysis described various changes in peak shapes, position and intensity throughout the fingerprint region of 1800-

900cm^{-1}, corresponding to a change in proteins, Amide I II and III, nucleic acids, lipid and carbohydrate. There was a significant contribution from paraffin despite efforts to negate this in the data analysis methods applied. Overall DNA and Amide III peaks were more dominant in malignant cells, indicating increased cellular content, consistent with the current literature findings. The dysplasia group proved the most difficult to analyse as it spanned the breadth of findings and was really very varied. This was reflected in the PCA models, which detailed the spread with a wide- tailed scatter plot (see Figure 5.23). This probably reflected the inter-group pathological variability, ranging from close-to-normal to close-to-cancer, and hence why it is in keeping with the expectation that dysplastic cells by the nature of their biochemistry are varied, and despite trying to reduce this by only using 'low grade' or 'high grade' can still show huge biochemical differences across the group.

Various methods were introduced in order to optimise the PCA-fed LDA classification models, including application of standard deviation to filter outliers and normalising data, however the Four Group Model demonstrated a high average specificity (75- 89%) but low average sensitivity (41- 60%) following LOSOCV, which reflected a poor measure of PCA-fed LDA classification model performance. This was improved, although not dramatically, when the spectra were cut to 178,000 per group to reduce bias for unequal dataset. Other causes for the large number of misclassified spectra include problems with epithelial cell selection, inclusion of poor quality spectra, and inter-sample spectral variability within the groups.

The Two Group Model (EM compared to combined cancers) however demonstrated overall a much better performance with an average sensitivity of 75% and specificity of 82%. This model demonstrates a good ability to distinguish between the two pathology groups with accuracy and reliability. The improvement between the two Models may be because of the sheer numbers of spectra involved in the first model (over 1.5 million) may actually flood the model without being of use, there may be a saturation point which beyond, adding more spectra does not improve the discrimination model because population variance has been captured. The fact that the Four Group Model improved with spectral reduction is in keeping with this, and would certainly be an area to examine in more detail in the future.

Several other major factors were discovered which might influence the performance of the discrimination models based on the experimental design;

Firstly the pathological methods must be scrutinised for weakness. All samples were analysed by two different consultants for the diagnosis, however only one identified the glands for the experimentation, and it was hard to identify the same gland in some cases, as the plane of examination changed with the changing depth of every section through the specimen. These glands were not always in the same stage of the cell cycle, or even the same orientation, therefore the spectral analysis may have been at varying depths in the crypts thus introducing variability in the genetic biochemistry (proliferation occurs at the base of crypts and differentiation variably higher up the crypts). Once the experiments were underway I again found it difficult to identify exactly the same glands that had been identified in the initial white light diagrams, and, despite the use of detailed topographical maps, sometimes it was simply impossible to identify the exact same ones which would have introduced further variability.

There also appeared to be some evidence of a few times where normal glands were present amongst pathological glands (particularly in the Dysplasia group) and so this would have contributed to the variability. I would be keen to address this further with a detailed analysis of the spectra selected and compared with the classification model, and then go back to the pathologists to re-examine the slides. This would be difficult as some of the white light images were poor, and especially in the Cancer groups, were difficult to interpret. Figure 6.1 demonstrates an intensity image of a cancer gland with little or no identifiable landmark and shows just how difficult epithelial selection was in some cases, compounded on occasion, by cautery artefact from the polypectomy procedure damaging tissue quality resulting in the loss of tissue landmarks, such as seems evident on this image.
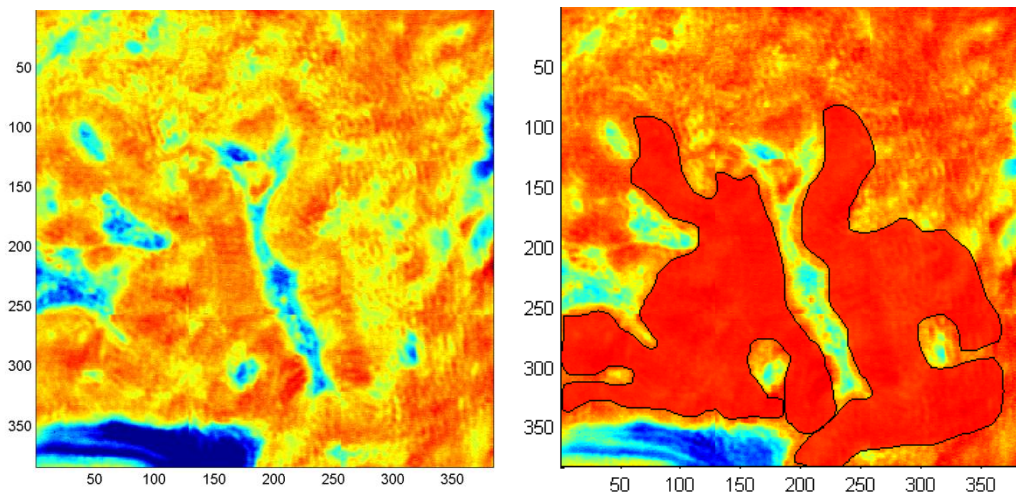
Figure 6.1: (Above) The white light image for a cancer sample and (Below left): The corresponding intensity image with limited glandular landmark for epithelial selection, as seen on the right

There were various problems with the instrumentation and set-up, compounded because the use of the high resolution technique is really still only early in its infancy. The study design was aided by work from the MiNERVA project (see Chapter 3), but often decisions made about set-up were being learnt as the experiments proceeded. For example, the number of acquisitions per sample was set at 64 (background 256) which was to keep the time per sample at a reasonable length. However increasing this would have improved spectral

quality and this should be investigated in the future. I would also change the background acquisition in future experiments. It was decided to perform a background every three specimens because there was not always enough clear room on each slide to take a measurement. However, it became clear that the spectrometer warmed significantly between the first and third specimen and so one background was not sufficient over this time. Ideally a background should be performed every single time so that variation in CO_2, water and heat could have been made. I made this choice because of time and space on the slide pressures but in the future would most certainly change this aspect of data collection, including switching to bigger calcium fluoride slides if the polyp was large and more room was needed to give a clear area around it.

Other causes of reduced signal to noise ratio included a centre-burst problem which caused the results to be affected in one tile per set. There was also no stage calibration on the spectrometer so I was unable to come back to the exact same image when interchanging swinging between white light and infrared and back again.

There was initially a problem in calibration between white light and infrared, so much so that when the image had been identified in white light the infrared moved a whole tile 'shot' away, so this had to be taken into account when choosing a gland and definitely affected the quality of the gland selected and then the ability to come back at a later date and select the epithelium, because the white light and infrared image were completely misaligned. This was corrected by Agilent technicians but only after a significant number of experiments had already been conducted.

I also became concerned with the quality of the transportation and movement of slides in the pathology cardboard trays to and from Exeter over the course of the experiments. It became obvious that fluff and debris from the cardboard trays were contaminating the slides. They were also getting scratched and showing considerable deterioration in condition of the slides. This would have caused increased scatter and artefact to be introduced into spectra and therefore reduced the overall quality of the spectra obtained. To combat this problem slides should be stored in a specific rigid plastic container to prevent

them knocking and moving during transport which would minimise the wear and tear over time.

7.1 Future Work

As well as making the changes suggested in experimental design as previously indicated, optimisation of the PCA-fed LDA classification model would be an important focus of future work to improve the accurate classification of colonic pathologies. Firstly selection of the spectra should be more tightly controlled. This should be investigated in a variety of ways but initially a case-specific selection would be implemented, by which the individual classification for each case would be analysed separately and only included if the values reached a certain percentage (for example, if a case had 20% Normal, 50% Dysplasia and 30% EM classified spectra it would be classified overall as a dysplasia sample). Analysis of the Dysplasia group should be performed in much greater detail, potentially reducing the number of outlying spectra in the wide 'tail' on the scatter plot. The classified spectra could be plotted over the top of the original selected spectra on the intensity map and then outliers could be identified, areas where misclassification was performed by human error would be identified, and selection of spectra could be reassessed and hopefully improved. Further improvement of the models could be made by changing the number of principal components (PCs) used, more careful selection of PCs and more variation in the LDA. The standard deviation filter could be extended and moderated to exclude more of the outlying spectra too, which would improve the performance of the model.

Finally the model should be tested with new samples rather than using the leave one sample out cross validation techniques. LOSOCV is an excellent alternative when there are no new samples, but to rigorously test the prediction model, new samples should be used which are completely unknown to it. This would give a completely unbiased and totally independent and more robust reflection of the performance of the model along with average sensitivity and specificity.

In the future the number of specimens, with which to train the model and increase population variance, should be increased. This should also increase

the breadth of pathology, potentially to include other benign pathologies such as inflammatory bowel disease as these are areas of clinical need and would be complimentary to Raman spectroscopy use in the colorectum in real-time in difficult diagnostic areas. Larger spectral models are required to reflect the great variation expected to be encountered clinically.

Other forms of chemometric analysis could also be explored, for example the use of artificial neural networks and cluster analysis to extend the predictive capacity and therefore usefulness of the model. These could be implemented also to analyse the spectral quality of these results and streamline the model further.

Exploitation of other optical analysis techniques could be used as a complimentary technique for further biochemical analysis of colon cases. This would necessitate prospective data collection so that samples could be fresh frozen, for example for Raman spectroscopy, as well as paraffin embedded for the infrared spectroscopy, but this would generate a great deal more scope in data analysis across the dataset.

Overall, this study has demonstrated the use of high resolution infrared spectroscopy as a complimentary diagnostic tool in colonic pathology analysis, with its specific potential use in the differentiation of epithelial misplacement from cancer. This is an area where early detection and diagnosis is key, both to ongoing patient management and the time taken to diagnosis, which is extremely important in current NHS practice. (NHS Interim Management and Support 2014) There is still much work before it can be clinically implemented, including a large multicentre trial to evaluate its use further. High resolution spectroscopy is certainly an exciting area of optical analysis and one in which routinely processed samples can now be analysed and yield a huge amount more data than ever before with standard infrared spectroscopy, thus increasing its clinical and commercial potential in line with that of Raman techniques and the in-vivo real time analysis of samples. The next step in the use of high resolution techniques will have to deal with the huge volume of data and computer analysis, and be responsive to the risk that an experiment might

potentially generate more data than is actually useful or can be analysed adequately.

APPENDIX A

Mean Spectra for all groups (except Normal group which remained unchanged) following manual exclusion of spectra.

Mean Spectra for All Dysplasia Cases Post manual Exclusion
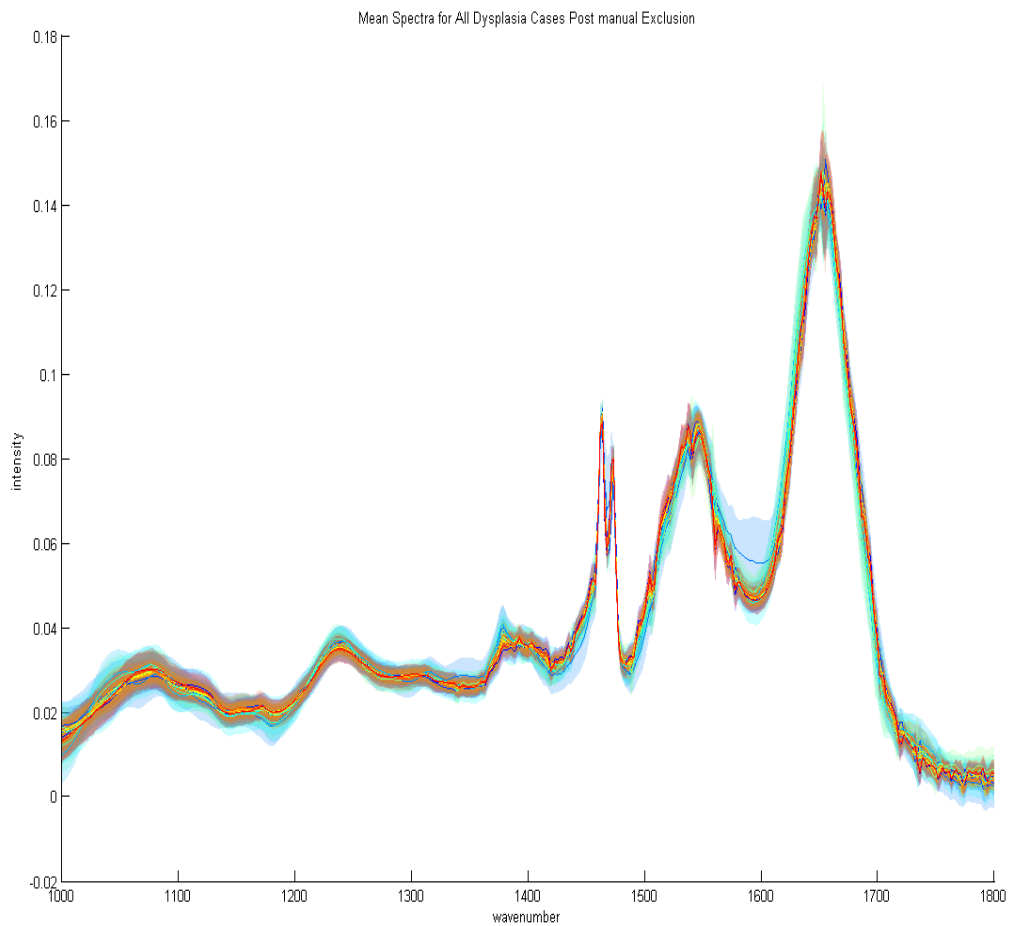
Figure 1.1: Mean spectra for all cases plotted together for Dysplasia group post manual exclusion

Mean Spectra for All Cancer Cases Post Manual Exclusion
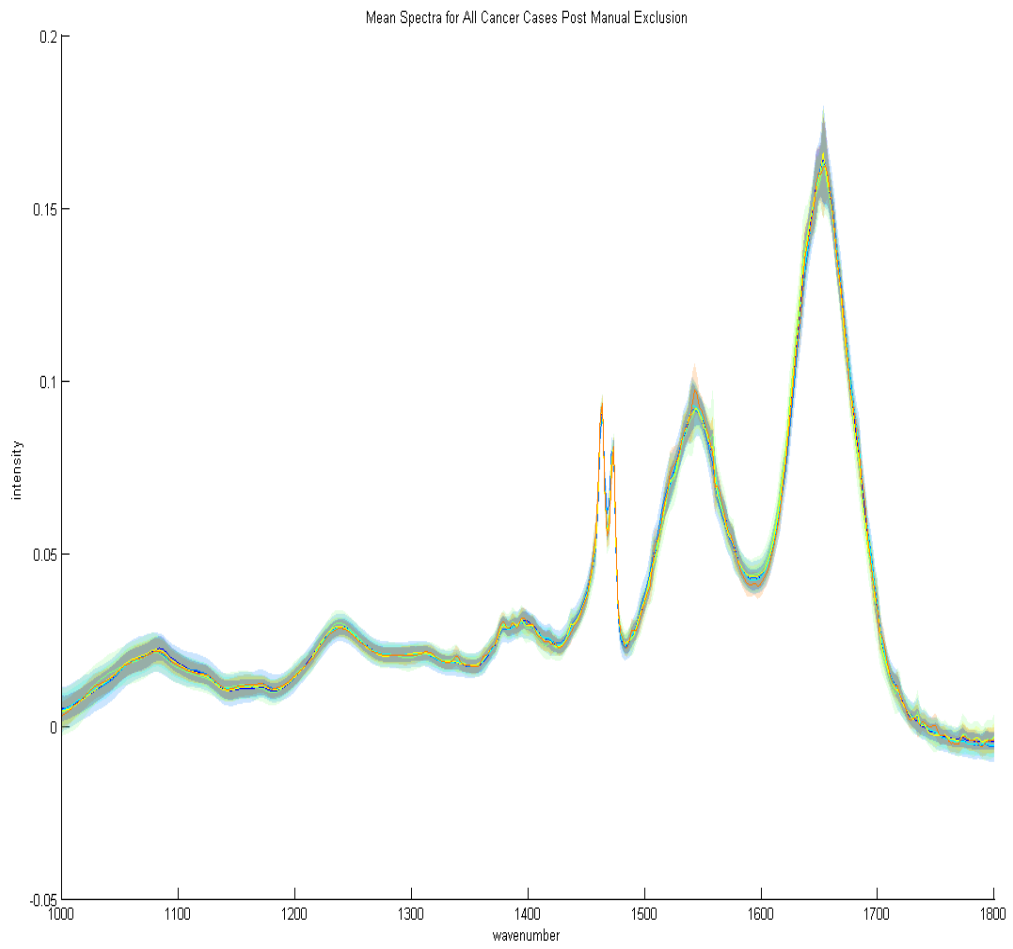
Figure 1.2: Mean spectra for all cases plotted together for Cancer group post manual exclusion

Mean Spectra for All Polyp Cancer Cases Post Manual Exclusion
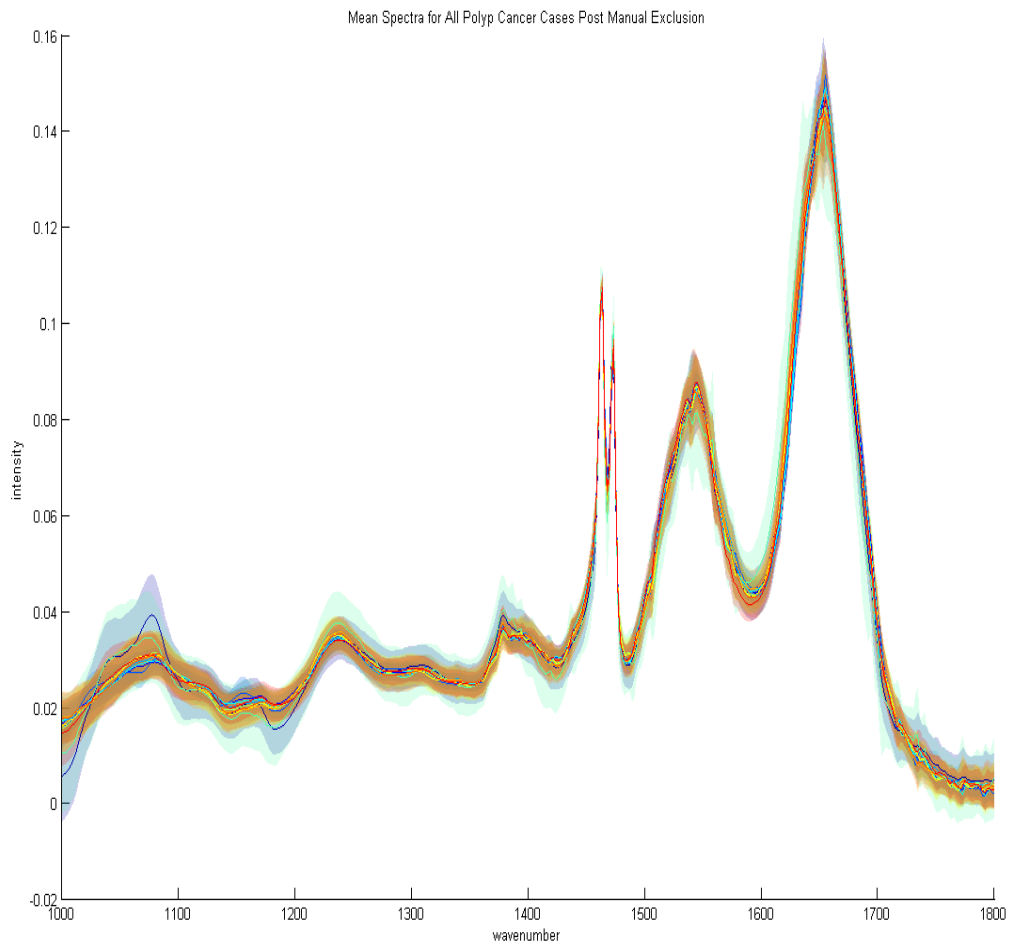
Figure 1.3: Mean spectra for all cases plotted together for Polyp Cancer group post manual exclusion

Mean Spectra for All Epithelial Misplacement Cases Post Manual Exclusion
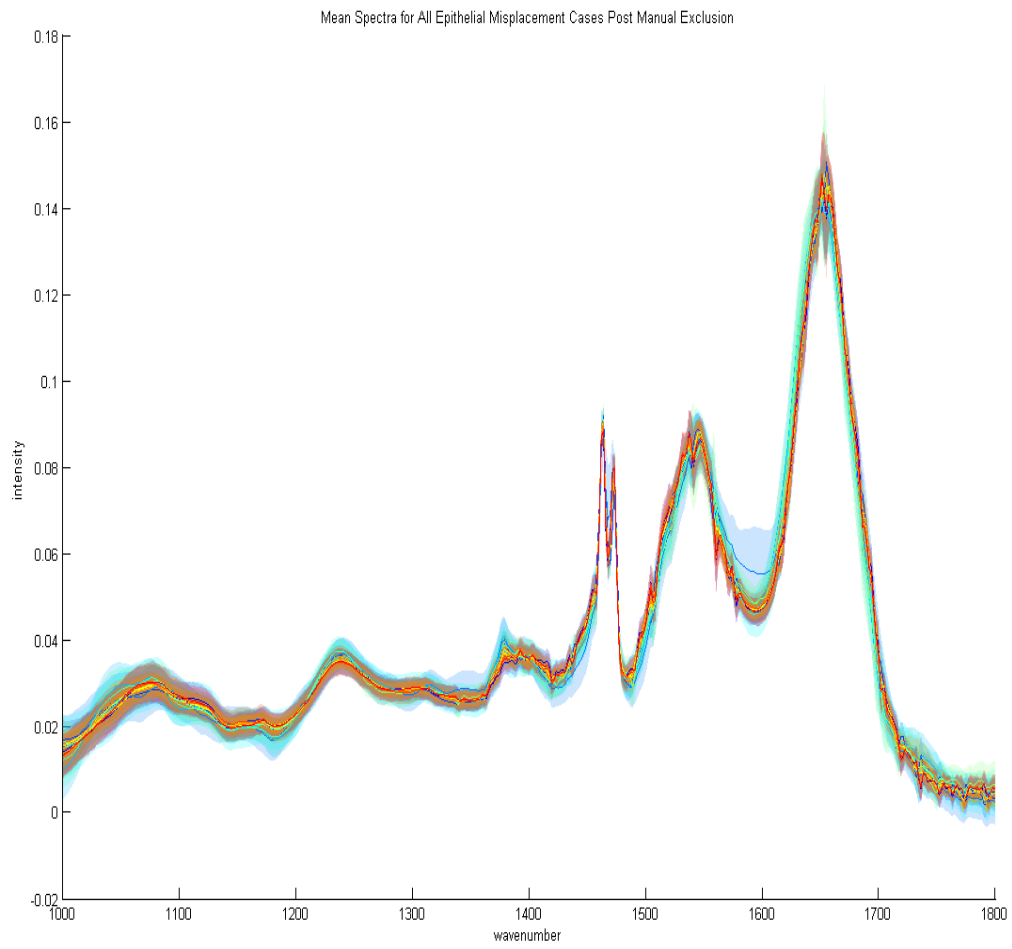
Figure 1.4: Mean spectra for EM group post manual exclusion

BIBLIOGRAPHY

Almond LM (2012) Assessment of a custom-built Raman spectroscopic probe for diagnosis of early oesophageal neoplasia. J Biomed Opt 17:081421–1 – 6.

Ami D, Bonecchi L, Cali S, Orsini G, Tonon G, Doglia SM (2003) FT—IR study of heterologous protein expression in recombinant Escherichia coli strains. Biochim Biophys Acta 1624:6–10.

Andersson-Engels S, Johansson J, Svanberg K, Svanberg S (1990) Laser-induced fluorescence in medical diagnostics. Proc SPIE 1203:76–96.

Andrus P, Strickland R (1998) Cancer grading by Fourier transform infrared spectroscopy. Biospectroscopy 4:37–46.

Argov S, Ramesh J, Salman A, Sinelnikov I, Goldstein J, Guterman H, Mordechai S (2002) Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients. J Biomed Opt 7:248–54.

Ashok PC, Praveen BB, Bellini N, Riches A, Dholakia K, Herrington CS (2013) Multi-modal approach using Raman spectroscopy and optical coherence tomography for the discrimination of colonic adenocarcinoma from normal colon. Biomed Opt Express 4:2179–86.

Atkin WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JMA, Parkin DM, Wardle J, Duffy SW, Cuzick J (2010) Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 375:1624–33.

Atreya R, Neurath MF (2008) Signaling molecules: the pathogenic role of the IL-6/STAT-3 trans signaling pathway in intestinal inflammation and in colonic cancer. Curr Drug Targets 9:369–74.

Barth A (2007) Infrared spectroscopy of proteins. Biochim Biophys Acta 1767:1073–101.

Bartholomew LG, Dahlin DC, Waugh JM (1957) Intestinal polyposis associated with mucocutaneous melanin pigmentation Peutz-Jeghers syndrome; review of literature and report of six cases with special reference to pathologic findings. Gastroenterology 32:434–51.

BBC Health (2014) Goggles help surgeons "see" tumours. http://www.bbc.co.uk/news/health-26954138. Accessed 10 Oct 2014

Bellisola G, Sorio C (2012) Infrared spectroscopy and microscopy in cancer research

and diagnosis. Am J Cancer Res 2:1–21.

Bird B, Miljkovic M, Romeo M, Smith J, Stone N, George M, Diem M (2008) Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology.

Bujanda L (2010) Malignant colorectal polyps. World J Gastroenterol 16:3103.

Cancer Research UK (2010) Cancer Stats. www.cancerresearchuk.org/cancerstats.

Chan JW, Taylor DS, Zwerdling T, Lane SM, Lhara K, Huser T (2006) Micro-Raman spectroscopy detects individual neoplastic and normal haematopoietic cells. J Biophys 90:648–656.

Chiriboga L, Xie P, Yee H, Vigorita V, Zarou D, Zakim D, Diem M (1998) Infrared spectroscopy of human tissue. Differentiation and maturation of epithelial cells in the human cervix. Biospectroscopy 4:47–53.

Clouston AD, Walker NI (2013) Polyps and tumour-like lesions of the large intestine. In: Shepherd NA, Warren BF, Williams GT, Greenson JK, Lauwers GY, Novelli MR (eds) Morson and Dawson's Gastrointestinal Pathology. pp 652–653

CNX (2014) Histology of the Alimentary Canal. In: Overv. Dig. Syst. http://cnx.org/contents/94fbc1ca-dcaa-47ad-94be-2cfeb168c347@3/Overview-of-the-Digestive-Syst. Accessed 11 Mar 2015

Cohenford M a, Lim S, Brown C, Chaudhry M a, Sigdel S, Beckelhimer E, Rigas B (2012) FT-IR microspectroscopy of mouse colon tissues: insight into the chemistry of carcinogenesis and diagnostic potential. Am J Pathol 181:1961–8.

Conti C, Ferraris P, Giorgini E, Rubini C, Sabbatini S, Tosi G, Anastassopoulou J, Arapantoni P, Boukaki E, Konstadoudakis S, Theophanides T, Valavanis C (2008) FT-IR microimaging spectroscopy: A comparison between healthy and neoplastic human colon tissues. J Mol Struct 881:46–51.

Davis A (2007) Light-Dark FTIR Absorbance Difference Spectroscopy for the Study of Photosystem I. Brock University

Dhar A, Johnson KS, Novelli MR, Bown SG, Bigio IJ, Lovat LB, Bloom SL (2006) Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique. Gastrointest Endosc 63:257–261.

Dovbeshko G, Chegel V, Gridina N, Repnytska O, Shirshov Y, Tryndiak V, Todor I, Solyanik G (2002) Surface enhanced IR absorption of nucleic acids from tumour cells: FTIR reflectance study. Biopolym 67:470–486.

Dovbeshko G, Gridina N, Kruglova E, Pashchuk O (1997) FTIR spectroscopy studies of nucleic acid damage. Talanta 53:233–246.

Eaden JA, Abrams KR, Mayberry JF (2001) The risk of colorectal cancer in ulcerative colitis: a meta-analysis. Gut 48:526–35.

Eye Site Texas (2015) Optical Coherence Tomography. Painless, Non-Invasive Imaging. http://www.eyesitetexas.com/optical-coherence-tomography.aspx. Accessed 25 Mar 2015

Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61:759–67.

Findlay CRJ, Wiens RA, Rak M, Sedlmair J, Hirschmugl C, Morrison J, Mundy CJ, Kansiz M, Gough KM (2015) Rapid biodiagnostic ex vivo imaging at 1 μm pixel resolution with thermal source FTIR FPA. Analyst 140:2493–503.

Fraser C (2009) Tailoring treatment according to the cancer genes. https://i0.wp.com/www.londongastroenterologist.co/images/large/Polyp-to-bowel-cancer-sequence.jpg. Accessed 13 Mar 2015

Fujimoto J, Brezinski M, Tearney G, Boppart S, Bouma B, Hee M, Southern J, Swanson E (1995) Optical biopsy and imaging using optical coherence tomography. Nat Med 1:970–972.

Fukuyama Y, Yoshida S, Yanagisawa S, Shimizu M (1999) A study on the differences between oral squamous cell carcinomas and normal oral mucosas measured by Fourier transform infrared spectroscopy. Biospectroscopy 5:117–126.

Fung M, Senterman M, Mikhael N, Lacelle S, Wong P (1996) Pressure-tuning Fourier transform infrared spectroscopic study of carcinogenesis in human endometrium. Biospectroscopy 2:155–165.

Gasparri F, Muzio M (2003) Monitoring of apoptosis of HL60 cells by Fourier-transform infrared spectroscopy. Biochem J 369:239–248.

Genericlook.com (2010) The Sections Of The Large Intestine. http://medicalterms.info/anatomy/Large-Intestine/. Accessed 10 Mar 2015

Gossage K, Tkaczyk T, Rodriguez J, Barton J (2003) Texture analysis of optical coherence tomography images: feasibility for tissue classification. J Biomed Opt 8:570– 575.

Hafiane A, Bunyak F, Palaniappan K (2008) Fuzzy Clustering And Active Contours for Histopathology Image Segmentation and Nuclei Detection. In: Lecture Notes in Computer Sciences. pp 903–914

Haggitt RC, Glotzbach RE, Soffer EE, Wruble LD (1985) Prognostic factors in colorectal carcinomas arising in adenomas: implications for lesions removed by endoscopic polypectomy. Gastroenterology 89:328–36.

Hansen TP, Fenger C, Kronborg O (1999) The Expression of p53, Ki-67 and Urokinase

Plasminogen Activator Receptor in Colorectal Adenomas with True Invasion and Pseudoinvasion. APMIS 107:689–94.

Herzberg G (1950) Rotation-vibration spectrum of the HD molecule. Nature 166:563.

Hewitson P, Glasziou P, Watson E, Towler B, Irwig L (2008) Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 103:1541–1549.

Horiuchi A, Tanaka N (2014) Improving quality measures in colonoscopy and its therapeutic intervention. World J Gastroenterol 20:13027–13034.

Hsing a W, McLaughlin JK, Chow WH, Schuman LM, Co Chien HT, Gridley G, Bjelke E, Wacholder S, Blot WJ (1998) Risk factors for colorectal cancer in a prospective study among U.S. white men. Int J cancer 77:549–53.

Huleihel M, Salman A, Erukhimovich V, Ramesh J, Hammody Z, Mordechai S (2002) Novel optical method for study of viral carcinogenesis in vitro. J Biochem Biophys Methods 50:111–121.

Hurley J, Morgan M, Williams G, Dolwani S (2013) Histopathological uncertainties in the management of early colorectal cancers resected through endoscopic therapy. Gut 62:A146–A147.

Inomata H, Tamai N, Aihara H, Sumiyama K, Saito S, Kato T, Tajiri H (2013) Efficacy of a novel auto-fluorescence imaging system with computer-assisted color analysis for assessment of colorectal lesions. World J Gastroenterol 19:7146–53.

Isabelle M, Stone N, Barr H, Vipond M, Shepherd N, Rogers K (2008) Lymph node pathology using optical spectroscopy in cancer diagnostics. 22:97–104.

Kallaway C, Almond LM, Barr H, Wood J, Hutchings J, Kendall C, Stone N (2013) Advances in the clinical application of Raman spectroscopy for cancer diagnostics. Photodiagnosis Photodyn Ther 10:207–219.

Kallenbach-Thieltges A, Großerüschkamp F, Mosig A, Diem M, Tannapfel A, Gerwert K (2013) Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue sections. J Biophotonics 6:88–100.

Kelloff GJ, Schilsky RL, Alberts DS, Day RW, Guyton KZ, Pearce HL, Peck JC, Phillips R, Sigman CC (2004) Colorectal adenomas: a prototype for the use of surrogate end points in the development of cancer prevention drugs. Clin Cancer Res 10:3908–18.

Kendall C, Isabelle M, BazantF B-H, Hutchings J, Orr LE, Babrah J, Baker R, Stone PN (2009) Vibrational spectroscopy: a clinical tool for cancer diagnostics. Analyst 134:1029–1045.

Kiesslich R, Burg J, Vieth M, Gnaendiger J, Enders M, Delaney P, Polglase A, McLaren

W, Janell D, Thomas S, Nafe B, Galle P, Neurath M (2004) Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo. Gastroenterology 127:706– 713.

Krafft C, Codrich D, Pelizzo G, Sergo V (2008) Raman and FTIR microscopic imaging of colon tissue: a comparative study. J Biophotonics 1:154–69.

Krafft C, Steiner G, Beleites C, Salzer R (2009) Disease recognition by infrared and Raman spectroscopy. J Biophotonics 2:13–28.

Kudo S, Hirota S, Nakajima T, Hosobe S, Kusaka H, Kobayashi T, Himori M, Yagyuu A (1994) Colorectal tumours and pit pattern. J Clin Pathol 47:880–885.

Kuiper T, van den Broek FJC, Naber AH, van Soest EJ, Scholten P, Mallant-Hent RC, van den Brande J, Jansen JM, van Oijen AH a M, Marsman W a, Bergman JJGHM, Fockens P, Dekker E (2011) Endoscopic trimodal imaging detects colonic neoplasia as well as standard video endoscopy. Gastroenterology 140:1887–94.

Lasch P, Diem M, Hansch W, Naumann D (2006) Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging. J Chemom 20:209– 220.

Lasch P, Haensch W, Naumann D, Diem M (2004) Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. Biochim Biophys Acta 1688:176–86.

Le Naour F, Bralet M, Debois D, Sandt C, Guettier C, Dumas P, Brunelle A, Laprévote O (2009) Chemical imaging on liver steatosis using synchrotron infrared and ToF-SIMS microspectroscopies. PLoS One 4:e7408.

Li Q-B, Xu Z, Zhang N-W, Zhang L, Wang F, Yang L-M, Wang J-S, Zhou S, Zhang Y-F, Zhou X-S, Shi J-S, Wu J-G (2005) In vivo and in situ detection of colorectal cancer using Fourier transform infrared spectroscopy. World J Gastroenterol 11:327–30.

Lichtenstein P, Holm N V, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78–85.

Lieberman DE, Weiss DG, Bond JH, Ahnen DJ, Garewal H, Chejfec G (2000) Use of Colonoscopy to Screen Asymptomatic Adults for Colorectal Cancer. N Engl J Med 343:162–168.

Liu HH, Kudo SE, Juch JP (2003) Pit pattern analysis by magnifying chromoendoscopy for the diagnosis of colorectal polyps. J Formos Med Assoc 102:178–82.

Liu Y, Njuguna R, Matthews T, Akers WJ, Sudlow GP, Mondal S, Tang R, Gruev V, Achilefu S (2013) Near-infrared fluorescence goggle system with complementary

metal – oxide – semiconductor imaging sensor and see- through display. J Biomed Opt 18:101303–1 – 101303–7.

Lockhart-Mummery J (1928) Diagnosis of Diseases of the Rectum and Colon. Postgrad Med J 3:96–100.

Loughrey MB, Quirke P, Shepherd NA (2014) Standards and datasets for reporting cancers Dataset for colorectal cancer histopathology reports July 2014.

Loughrey MB, Shepherd NA (2015) The pathology of bowel cancer screening. Histopathology 66:66–77.

Lui K., Li J, Kelsey SM, Mantsch HH (2001) Quantitative determination of apoptosis on leukaemia cells by infrared spectroscopy. Apoptosis 6:267–276.

Lyng FM, Ramos IRM, Ibrahim O, Byrne HJ (2015) Vibrational microspectroscopy for Cancer Screening. Applies Sci 5:23–35.

Martin FL, Kelly JG, Llabjani V, Martin-Hirsch PL, Patel II, Trevisan J, Fullwood NJ, Walsh MJ (2010) Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nat Protoc 5:1748–60.

Mordechai S, Sahu R, Hammody Z, Mark S, Kantarovich K, Guterman H, Podshyvalov J, Goldstein J, Argov S (2004) Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma. J Microsc 215:86–91.

Mourant JR, Bigio IJ, Boyer J, Johnson TM, Lacey J, Bohorfoush AG, Mellow M (1996) Elastic Scattering Spectroscopy as a diagnostic tool for differentiating pathologies in the gastrointestinal tract: Preliminary testing. J Biomed Opt 1:192–199.

Movasaghi Z, Rehman S, ur Rehman DI (2008) Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. Appl Spectrosc Rev 43:134–179.

Mudter J, Neurath MF (2007) Il-6 signaling in inflammatory bowel disease: pathophysiological role and clinical relevance. Inflamm Bowel Dis 13:1016–23.

Mueller J, Mueller E, Arras E, Bethke B, Stolte M (1997) Stromelysin-3 Expression in Early (pT1) Carcinomas and Pseudoinvasive Lesions of the Colorectum. Virchows Arch 430:213:219.

Muto T, Bussey H, Morson B (1975) The evolution of cancer of the colon and rectum. Cancer 36:2251 – 2270.

Muto T, Bussey H, Morson B (1973) Pseudo-carcinomatous invasion in adenomatous polyps of colon and rectum. J Clin Pathol 26:25–31.

Nallala J, Gobinet C, Diebold M-D, Untereiner V, Bouché O, Manfait M, Sockalingum GD, Piot O (2012) Infrared spectral imaging as a novel approach for histopathological recognition in colon cancer diagnosis. J Biomed Opt 17:116013.

Nallala J, Piot O, Diebold M-D, Gobinet C, Bouché O, Manfait M, Sockalingum GD (2013) Infrared imaging as a cancer diagnostic tool: introducing a new concept of spectral barcodes for identifying molecular changes in colon tumors. Cytometry A 83:294–300.

Napier B (2012) Minerva Project. http://minerva-project.eu/. Accessed 23 Feb 2015

National Institute for Health and Care Excellence (2011) NICE Colonoscopy Surveillance Guidelines.

Naumann D (2008) Biomedical Vibrational Spectroscopy. In: Lasch P, Kneipp J (eds) Biomedical Vibrational Spectroscopy. Wiley Interscience, Hoboken, pp 1–8

NHS England (2011) NHS Cancer Outcomes Strategy. In: NHS Choices Natl. Serv. Framew. http://www.nhs.uk/choiceintheNHS/Rightsandpledges/Waitingtimes/Pages/Guide to waiting times.aspx. Accessed 12 Nov 2014

NHS Interim Management and Support (2014) Delivering Cancer Waiting Times- A Good Practice Guide.

Noffsinger AE (2009) Serrated polyps and colorectal cancer: new pathway to malignancy. Annu Rev Pathol 4:343–64.

Nusko G, Mansmann U, Altendorf-Hofmann A, Groitl H, Wittekind C, Hahn EG (1997) Risk of invasive carcinoma in colorectal adenomas assessed by size and site. Int J Colorectal Dis 12:267–71.

Paluszkiewicz C, Kwiatek W (2001) Analysis of human cancer prostate tissues using FTIR microscopy and SXIXE techniques. J Mol Struct 565-566:329–334.

Parsa S (2008) Anatomy, Histology and Embrology of the Large Bowel. https://anatomytopics.wordpress.com/2008/12/21/21-anatomy-histology-embryology-of-the-large-intestine/. Accessed 7 Aug 2014

Petersen VC, Sheehan AL, Bryan RL, Armstrong CP, Shepherd NA (2000) Misplacement of dysplastic epithelium in Peutz-Jeghers Polyps: the ultimate diagnostic pitfall? Am J Surg Pathol 24:34–9.

Pohl J, Schneider A, Vogell H, Mayer G, Kaiser G, Ell C (2011) Pancolonic chromoendoscopy with indigo carmine versus standard colonoscopy for detection of neoplastic lesions: a randomised two-centre trial. Gut 60:485–490.

Pritchard CC, Grady WM (2011) Colorectal cancer molecular biology moves into clinical practice. Gut 60:116–29.

Public Health England (2013) NHS Bowel Cancer Screening Programme. http://www.cancerscreening.nhs.uk/bowel/index.html.

Quaroni L, Casson A (2009) Characterization of Barrett esophagus and esophageal adenocarcinoma by Fourier-transform infrared microscopy. Analyst 134:1240–1246.

Ragazzi M, Piana S, Longo C, Castagnetti F, Foroni M, Ferrari G, Gardini G, Pellacani G (2014) Fluorescence confocal microscopy for pathologists. Mod Pathol 27:460–471.

Reddy RK, Walsh MJ, Schulmerich M V., Carney PS, Bhargava R (2013) High-definition infrared spectroscopic imaging. Appl Spectrosc 67:93–105.

Rehman S, Movasaghib Z, Darr J, Rehman I (2010) Fourier Transform Infrared Spectroscopic Analysis of Breast Cancer Tissues; Identifying Differences between Normal Breast, Invasive Ductal Carcinoma, and Ductal Carcinoma In Situ of the Breast. Appl Spectrosc Rev 45:355–368.

Richter T, Steiner G, Abu-Id M, Salzer R, Gergmann R, Rodig H, Johannsen B (2002) Identification of tumor tissue by FTIR spectroscopy in combination with positron emission tomography. Vib Spectrosc 28:103–110.

Riddell RH, Goldman H, Ransohoff DF, Appelman HD, Fenoglio CM, Haggitt RC, Ahren C, Correa P, Hamilton SR, Morson BC (1983) Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. Hum Pathol 14:931–68.

Rigas B, Morgello S, Goldman IS, Wong PT (1990) Human colorectal cancers display abnormal Fourier-transform infrared spectra. Proc Natl Acad Sci U S A 87:8140–4.

Romeo MJ, Boydson-White S, Matthaus C, Miljkovic M, Bird B, Chernenko T, Diem M (2008) Biomedical Vibrational Spectroscopy. In: Lasch P, Kneipp J (eds) Biomedical Vibrational Spectroscopy. Wiley-Blackwell, p 124

Roy A, Teare J, Rao S, Cunningham D (2012) Clinical review- Colorectal cancer. http://www.gponline.com/clinical-review-colorectal-cancer/cancer/bowel/article/1317800. Accessed 10 Mar 2015

Roy HK, Goldberg MJ, Bajaj S, Backman V (2011) Colonoscopic Optical Biopsy: Bridging Technological Advances to Clinical Practice. Gastroenterology 140:1863–1867.

Sano Y, Ikematsu H, Fu KI, Emura F, Katagiri A, Horimatsu T, Kaneko K, Soetikno R, Yoshida S (2009) Meshed capillary vessels by use of narrow-band imaging for differential diagnosis of small colorectal polyps. Gastrointest Endosc 69:278–83.

Scholefield JH (2000) ABC of colorectal cancer: screening. BMJ 321:1004–6.

Sehgal R, Sheahan K, O'Connell PR, Hanly AM, Martin ST, Winter DC (2014) Lynch Syndrome: An Updated Review. Genes (Basel) 5:497–507.

Shepherd N, Griggs R (2015) Bowel cancer screening-generated diagnostic conundrum

of the century: pseudo-invasion in sigmoid colonic polyps. Mod Pathol 28:S88–S94.

Shepherd NA (2002) Polyps of the Large Intestine. In: Haboubi N, Geboes K, Shepherd N, Talbot I (eds) Gastrointestinal Polyps. Cambridge University Press, pp 95–122

Shepherd NA, Bussey HJ, Jass JR (1987) Epithelial misplacement in Peutz-Jeghers polyps. A diagnostic pitfall. Am J Surg Pathol 11:743–9.

Sherman C (2009) Colorectal Cancer: Prevention vs. Detection. http://www.clinicaladvisor.com/colorectal-cancer-prevention-vs-detection/article/126715/. Accessed 10 Mar 2015

Shiovitz S, Copeland WK, Passarelli MN, Burnett-Hartman AN, Grady WM, Potter JD, Gallinger S, Buchanan DD, Rosty C, Win AK, Jenkins M, Thibodeau SN, Haile R, Baron JA, Marchand LL, Newcomb PA, Lindor NM (2014) Characterisation of Familial Colorectal Cancer Type X, Lynch syndrome, and non-familial colorectal cancer. Br J Cancer 111:598–602.

Sivak M, Kobayashi K, Izatt J, Rollins A, Ung-Runyawee R, Chak A, Wong R, Isenberg G, Willis J (2000) High-resolution endoscopic imaging of the GI tract using optical coherence tomography. Gastrointest Endosc 51:474– 479.

Sobin LH (1985) Inverted hyperplastic polyps of the colon. Am J Surg Pathol 9:265–72.

Song LMWK, Adler DG, Chand B, Conway JD, Croffie JMB, DiSario JA, Mishkin DS, Shah RJ, Somogy L, Tierney WM, Petersen BT (2007) Chromoendoscopy. Gastrointest Endosc 66:639–649.

Sreedhar H, Varma VK, Nguyen PL, Davidson B, Akkina S, Guzman G, Setty S, Kajdacsy-Balla A, Walsh MJ (2015) High-definition Fourier Transform Infrared (FT-IR) Spectroscopic Imaging of Human Tissue Sections towards Improving Pathology. J Vis Exp 95:52332.

Subramanian V, Mannath J, Hawkey CJ, Ragunath K (2011) High definition colonoscopy vs. standard video endoscopy for the detection of colonic polyps: a meta-analysis. Endoscopy 43:499–505.

Svanberg K, Andersson-Engels S, Baert L, Bak-Jensen E, Berg R, Brun A, Colleen S, Idvall I, D'Hallewin M, Ingrar C, Johansson J, Karlsson S, Lundgren R, Salford LG, Stenram U, Strombald L, Svanberg S, Wang I (1994) Tissue characterization in some clinical specialties utilizing laser- induced fluorescence. Proc SPIE 2135:2–15.

Tanaka S, Sano Y (2011) Aim to unify the narrow band imaging (NBI) magnifying classification for colorectal tumors: current status in Japan from a summary of the consensus symposium in the 79th Annual Meeting of the Japan Gastroenterological Endoscopy Society. Dig Endosc 23 Suppl 1:131–9.

Taylor JC, Kendall C a, Stone N, Cook T a (2007) Optical adjuncts for enhanced colonoscopic diagnosis. Br J Surg 94:6–16.

Terry MB, Neugut AI, Bostick RM, Sandler RS, Haile RW, Jacobson JS, Fenoglio-Preiser CM, Potter JD (2002) Risk factors for advanced colorectal adenomas: a pooled analysis. Cancer Epidemiol Biomarkers Prev 11:622–9.

The Earth Observatory (2014) Does the Earth Have an Iris Analogue? http://earthobservatory.nasa.gov/Features/Iris/. Accessed 25 Mar 2015

Turner JK, Williams GT, Morgan M, Wright M, Dolwani S (2013) Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists. Histopathology 62:916–924.

Vo-Dinh T, Panjehpour M, Overholt BF, Farris C, Buckley FP, Sneed R (1995) In vivo cancer diagnosis of the esophagus using differential normalized fluorescence (DNF) indices. Lasers Surg Med 16:41–47.

Vrettos JS, Affleck RP, Guo J, Spitznagel TM, Krishnamurthy R (2006) Application of Difference Spectroscopy to Biopharmaceutical Formulation Development. In: Am. Lab. http://www.americanlaboratory.com/914-Application-Notes/19683-Application-of-Difference-Spectroscopy-to-Biopharmaceutical-Formulation-Development/. Accessed 18 Feb 2016

Winawer S, Fletcher R, Miller L, Godlee F, Stolar M, Mulrow C, Woolf S, Glick S, Ganiats T, Bond J, Rosen L, Zapka J, Olsen S, Giardiello F, Sisk J, Antwerp R Van, Brown-Davis C, Marciniak D, Mayer R (1997) Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology 112:594–642.

Wong NACS, Hunt LP, Novelli MR, Shepherd NA, Warren BF (2009) Observer agreement in the diagnosis of serrated polyps of the large bowel. Histopathology 55:63–6.

Wood B, Quinn M, Burden F, McNaughton D (1996) An investigation into FTIR spectroscopy as a bio-diagnostic tool for cervical cancer. Biospectroscopy 2:143–153.

Wood JJ, Kendall C, Hutchings J, Lloyd GR, Stone N, Shepherd N, Day J, Cook T a (2014) Evaluation of a confocal Raman probe for pathological diagnosis during colonoscopy. Colorectal Dis 16:732–8.

Wood JJ, Kendall C, LLoyd GR, Shepherd N a., Cook T a., Stone N (2011) Infrared spectroscopy to estimate the gross biochemistry associated with different colorectal pathologies. In: Ramanujam N, Popp J (eds) Proc. of SPIE-OSA Biomedical Optics. pp 80870P–1 –80870P–9

Yantiss R, Bosenberg M, Antonioli D, Odze R (2002) Utility of MMP-1, p53, E-cadherin,

and collagen IV immunohistochemical stains in the differential diagnosis of adenomas with misplaced epithelium versus adenomas with invasive adenocarcinoma. Am J Surg Pathol 26:206–15.

Yantiss RK, Goldman H, Odze RD (2001) Hyperplastic polyp with epithelial misplacement (inverted hyperplastic polyp): a clinicopathologic and immunohistochemical study of 19 cases. Mod Pathol 14:869–75.

Yoshida S, Miyazaki M, Sakai K, Takeshita M, Yuasa S, Sato A, Kobayashi T, Watanabe S, Okuyama H (1997) Fourier transform infrared spectroscopic analysis of rat brain microsomal membranes modified by dietary fatty acids: possible correlation with altered learning behaviour. Biospectroscopy 3:281–290.

Zhang K, Civan J, Mukherjee S, Patel F, Yang H (2014) Genetic variations in colorectal cancer risk and clinical outcome. World J Gastroenterol 20:4167–4177.

Zwielly A, Mordechai S, Sinielnikov I, Salman A, Bogomolny E, Argov S (2010) Advanced statistical techniques applied to comprehensive FTIR spectra on human colonic tissues. Med Phys 37:1047–55.