

# Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

# Infrared Micro-spectroscopy for Cyto-pathological Classification of Esophageal Cells

Douglas Townsend<sup>1</sup>, Miloš Mijković<sup>1</sup>, Benjamin Bird<sup>1</sup>, Kathleen Lenau<sup>1</sup>, Oliver Old<sup>2</sup>, Max Almond<sup>2</sup>, Catherine Kendall<sup>2</sup>, Gavin Lloyd<sup>2</sup>, Neil Shepherd<sup>3</sup>, Hugh Barr<sup>4</sup>, Nick Stone<sup>5</sup> and Max Diem<sup>1</sup>

<sup>1</sup>Laboratory for Spectral Diagnosis, Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA

<sup>2</sup>Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucester, UK

<sup>3</sup>Department of Pathology, Gloucestershire Royal Hospital, Great Western Road, Gloucester, UK

<sup>4</sup>Department of Upper GI Surgery, Gloucestershire Royal Hospital, Great Western Road, Gloucester, UK

<sup>5</sup>Biomedical Physics, School of Physics, University of Exeter, Exeter EX4 4QL, UK

## ABSTRACT

We report results from a study utilizing infrared spectral cytopathology (SCP) to detect abnormalities in exfoliated esophageal cells. SCP has been developed over the past decade as an ancillary tool to classical cytopathology. In SCP, the biochemical composition of individual cells is probed by collecting infrared absorption spectra from each individual, unstained cell, and correlating the observed spectral patterns, and the variations therein, against classical diagnostic methods to obtain an objective, machine-based classification of cells.

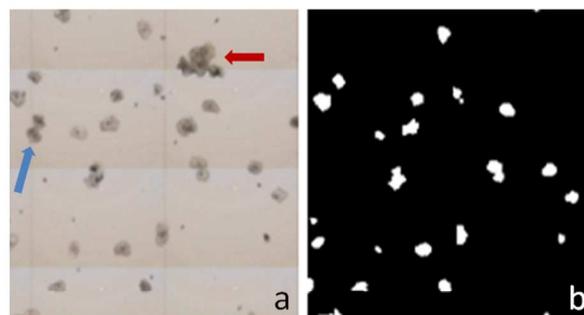
In the past, SCP has been applied to the analysis and classification of cells exfoliated from the cervix and the oral cavity. In these studies, it was established that SCP can distinguish normal and abnormal cell types. Furthermore, SCP can differentiate between truly normal cells, and cells with normal morphology from the vicinity of abnormalities. Thus, SCP may be a valuable tool for the screening of early stages of dysplasia and pre-cancer.

## INTRODUCTION

While overall cancer incidence in the US has decreased in recent years the number of new cases of esophageal cancer is increasing [1]. There are two major histological sub-types of esophageal cancer, squamous cell carcinoma and adenocarcinoma; the latter of which has seen an alarming increase in incidence in Western countries over the past several decades [2]. Early diagnosis of esophageal adenocarcinoma is of the utmost importance to allow curative intervention, as prognosis for patients with symptomatic and advanced adenocarcinoma is very poor. However late presentation is frequent, and overall 5-year mortality rates exceed 80%. Patients who are fit for radical surgical intervention have a median survival of 1 year and 5-year survival rates of 10% [3]; this makes esophageal adenocarcinoma one of the most lethal cancers.

Esophageal adenocarcinoma typically arises in the distal one-third portion of the esophagus and its primary risk factor is Barrett's esophagus; a metaplastic transformation of the esophageal mucosa that is thought to arise as a protective response to tissue inflammation as a result of chronic gastro-esophageal reflux disease (GERD) [4]. Barrett's esophagus is recognized by the transition from stratified squamous epithelial cells to specialized columnar cells and is present in about 15% of patients undergoing endoscopy for symptoms of GERD [5]. Although these columnar cells appear to be more resistant to acid reflux they are predisposed to carcinogenesis. With continued regurgitation of

gastric acid the columnar epithelium may become dysplastic, leading to the possibility of adenocarcinoma [6].



**Figure 1:** (a) White light image of a cellular region and (b) its corresponding binary mask generated in PapMap. Contiguous white areas correspond to cellular regions. (See text for details)

The increased risk of esophageal adenocarcinoma in patients with Barrett's esophagus has led to the development of endoscopic surveillance programs in many countries worldwide, recommending periodic upper-gastrointestinal endoscopy every 2-3 years to detect dysplastic change or early cancer [7].

Endoscopic surveillance is limited by the difficulty of detecting dysplasia or early cancer using white light endoscopy, with over 20% of lesions missed even by expert endoscopists using a high-resolution endoscope [8]. Consequently guidelines advocate random biopsy sampling of the mucosa

in 4 quadrants at 2 cm intervals throughout the region of Barrett's [9]. This results in lengthy endoscopic procedures and a high workload for pathologists.

The problem of accurate detections is further compounded by inter-observer variation in histopathology assessment. Inter- and intra- observer studies have demonstrated that, when dividing into 4 clinical categories (negative for dysplasia; indefinite for dysplasia and low-grade dysplasia; high-grade dysplasia; and adenocarcinoma), there are astonishingly poor levels of agreement among pathologists with intra-observer kappa values of 0.64; and inter-observer kappa values of 0.43 [10].

**Table I:** List of the original pathological diagnosis at the time of exfoliation for each sample in this study, as well as the number of individual cells from each respective pathological group.

Pathology	No. Samples	No. Cells studied
Normal Squamous	9	2339
Barrett's Esophagus	12	3197
Dysplasia	5	1161

Another major challenge in the early detection of esophageal adenocarcinoma is that less than 10% of cases arise in patients with [a prior diagnosis of](#) Barrett's esophagus [11]. This has led to consideration of screening programs to identify those with Barrett's esophagus at increased risk. However, endoscopic screening would be invasive and expensive, and consequently unlikely to be cost-effective [for a large](#) population. This problem has driven searches for non-endoscopic methods of detection of Barrett's. Esophageal cytology is a possible approach, with one such collection device (the cytosponge™) that has shown acceptability in a primary care setting [12]. If applied in clinical practice, assessment of esophageal cytology would require cytological expertise and consume significant pathology resources.

Infrared spectroscopy has been shown to discriminate between pathological cell types in a number of other disease states, such as oral, bladder and cervical cancer [13-15]. The current study investigated whether this application can be applied in the esophagus as a potential automated, objective diagnostic tool for detection of Barrett's esophagus.

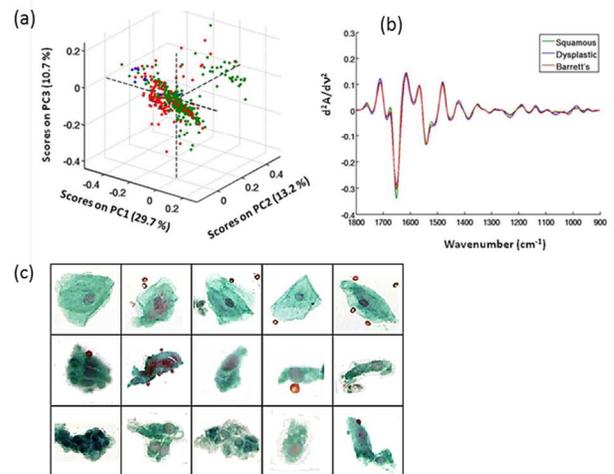
## MATERIALS AND METHODS

### Exfoliation

All cell samples for this study were collected during routine endoscopic or surgical procedures at the Department of Surgery at the Gloucestershire Royal Hospital, Gloucester, UK. The cell samples were exfoliated from regions of normal

squamous, Barrett's esophagus and dysplasia using a standard cytological brush attached to the endoscope. Samples on cytological brushes were fixed in 2% formalin solution and shipped to the Laboratory for Spectral Diagnosis at Northeastern University (Boston, MA, USA) for data acquisition. Cells were vortexed off the brushes, filtered to remove debris, and deposited onto 'low-e' microscope slides (Kevley Technologies, Chesterland, OH, USA) using cyto-centrifugation (CytoSpin, Thermo, Waltham, MA, USA).

Samples studied are described in Table I. Ethical approval for this study was provided by Gloucestershire Research Ethics Committee ReCAD1.

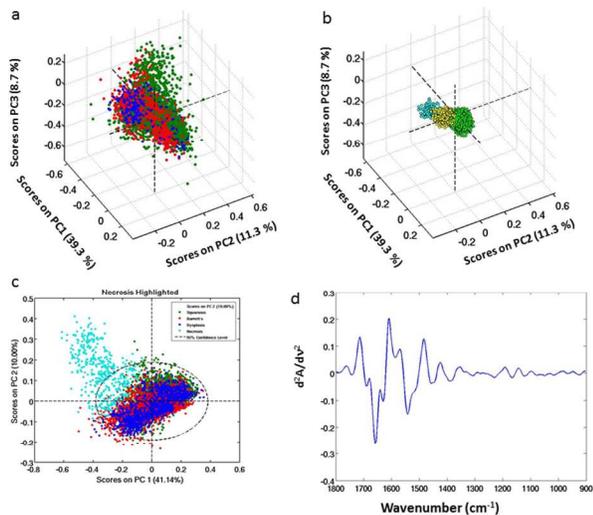


**Figure 2:** (a) PCA scores plot of diagnosed cells by a pathologist. Squamous are shown in green, Barrett's in red and dysplasia in blue. (b) Mean second derivative, vector normalized spectra showing distinct spectral differences between the disease states. (c) 40 x visual images of annotated cells some different pathological groups: (top) squamous, (middle) Barrett's and (bottom) dysplasia.

### Data Collection

Infrared hyperspectral data sets were collected from a 4 mm x 4 mm area on low-e slides in imaging mode using a Perkin Elmer Spectrum One/Spotlight 400 imaging IR microspectrometer (Sheldon, CT, USA) in the Laboratory for Spectral Diagnosis. The problems reported with the use of the low-e sample substrates [16] were accounted for by methods reported in the literature [17, 18]. The instrument bench, the IR microscope, and an external microscope enclosure box were purged with a continuous stream of dry air ( $-40\text{ }^{\circ}\text{C}$  dew point) to reduce atmospheric water vapor spectral contributions in the spectra. Data were acquired in reflectance mode with the following parameters:  $4\text{ cm}^{-1}$  spectral resolution,  $6.25\text{ }\mu\text{m}^2$  pixel size, 2 scans/pixel, Norton-Beer apodization, and 1 level of zero filling. The co-added interferograms for each pixel were Fourier transformed to yield spectral vectors, each with a range of  $4000\text{-}700\text{ cm}^{-1}$  at

2 cm<sup>-1</sup> intervals. Background spectra for all 16 detector ele-



**Figure 3:** (a) PCA scores plot of all spectra. (b) Same plot as (a) however the cells not within 1% of the mean annotated spectra are invisible. Squamous are shown in green, Barrett's in yellow and dysplastic in light blue. (c) PCA scores plot of spectra in (a) with cells falling within 5% of the mean necrosis vector, shown in (d), highlighted in light blue in panel c.

ments were collected using 128 co-added interferograms. Raw data sets consist of 409,600 spectra and occupy approximately 2.54GB each.

### Image Processing

Raw data sets were imported into software written in-house referred to as PapMap [19]. This program is written in 64-bit MATLAB (Mathworks, Natick, MA, USA) to accommodate the large data matrices. PapMap reconstructs the spectra of individual cells collected in mapping mode from between 10 and 100 individual pixel spectra (corresponding to 400 and 4000 μm<sup>2</sup>) for each raw cell area. To this end, PapMap first establishes which spectra belong to a given cell in the IR image map by constructing a binary mask in which contiguous regions belonging to individual cells are identified. This mask is based on raw amide I peak intensities that generally vary between 0.01 OD units for the cytoplasm to more than 0.1 OD units in the nucleus of a dried cell. Such a mask is shown in Figure 1b in which white regions correspond to cellular regions. There is a good correspondence between these regions and the visible image of the sample shown in Figure 1a; furthermore, the elimination of cell clumps (red arrow) and overlapping cells (blue arrow) is demonstrated by the fact that these regions are absent in the binary mask. For each contiguous area occupied by a cell, the cellular spectrum is calculated, starting from the spectrum with the largest amide I intensity. This spectrum is assumed to be

from the nucleus of the cell, which always exhibits the strongest protein intensity.

All spectra within the areas defined by the binary mask are subsequently co-added and subject to several constraints to eliminate spectra from the edges of the cell that may exhibit weak spectra with poor signal-to-noise, or spectra confounded by scattering artifacts [20]. The latter spectra were identified by lowered amide I frequencies. Subsequently, PapMap averages all spectra within the masked region to obtain one spectrum per cell. Cellular coordinates and nametags of each cell are also recorded so that correlation with standard pathology is possible. Subsequently, spectra are exported for further data analysis.

### Staining

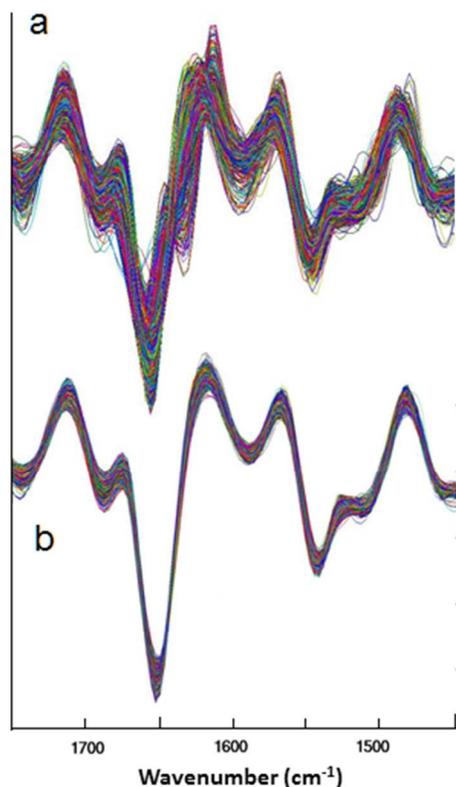
After IR data collection, the cells on a slide were manually stained using standard cytological stain combinations, as described in the literature [21]. Tap water and solutions of ethanol were used for the washing steps. Finally, to avoid degradation, slides are dipped in xylene and covered-slipped for cytological analysis. Using the cellular coordinates and nametags provided by PapMap, cell images are obtained using an Olympus BX40 microscope fitted with a computer-controlled stage and a QImaging GO3 3MB digital color camera. The images and cellular spectra are linked together and stored in a database for easy identification. Cellular images from 10 samples (8 Barrett's, 2 dysplastic) were analyzed by a pathologist at the Gloucestershire Royal Hospital.

### Data Analysis

Data analysis was performed using both supervised and unsupervised methods of multivariate analysis. In the latter category, principal component analysis (PCA) was used to determine whether or not there are distinct spectral differences between cellular spectra. The principal components (PCs) are obtained from the eigenvectors of the correlation matrix of the data set, and represent a completely unbiased decomposition of cellular spectra. PCA was carried out using the PLS Toolbox 402 in MATLAB (Eigenvector Research, Wenatchee, WA, USA) on the spectral range 1800-900 cm<sup>-1</sup> after second derivative calculation and subsequent vector normalization in this spectral range. This method is not suitable for diagnostic purposes since it merely displays the variance within a dataset, therefore supervised methods of analysis are required to predict class membership.

Artificial neural networks (ANNs) were used as supervised algorithms for diagnostic data analysis. The spectral database was split into independent training and testing sets by randomly selecting 400 spectra from each class for the training. Training and internal validation are carried out on the training spectra with known class assignments. Fifty

(intensity) features were selected in each spectral vector via entropy-based feature selection, using 'rankfeature' function in MATLAB. The MATLAB 'feed-forward' ANN function was employed for ANN construction, utilizing 50 inputs, one hidden layer with two nodes and one output neuron (binary classification of normal/abnormal). The ANN 'learns' to correlate spectral features of the training set with the diagnostic outcome by assigning 'weights' to each of the features [22]. External validation (testing) is carried out on equal numbers of spectra from each class. ANN performance is described in terms of sensitivity, specificity and overall accuracy.



**Figure 4:** Raw second derivative, vector normalized squamous spectra displaying the necrosis signal at  $1635\text{ cm}^{-1}$  (top) and filtered squamous cell spectra (bottom).

## RESULTS

We report SCP results from a relatively large annotated data set of individual cellular spectra: 299 squamous, 145 Barrett's and 14 dysplastic cells, totaling 458 diagnosed cells from 10 samples. The most abundant cell type found in each sample and disease class are the squamous cells of the esophagus, which are more easily exfoliated than the columnar cells.

The PCA 'scores' plot in Figure 2a shows the annotated set of spectra with squamous (green), Barrett's (red) and dysplasia (blue). In such a scores plot, each dot represents the spectrum of one cell, plotted in a coordinate system that indicates the contribution of each PC to a given spectrum. The PCs are 'basis' spectra calculated from the covariance within a data set and give indication of where significant biochemical differences occur in the spectra of cells [23]. In this scores plot, all the squamous cells diagnosed were from samples with disease. The cluster of squamous cells separating along PC2 was found to be cells from a dysplastic sample. Although the pathologist diagnosed these cells as healthy squamous their spectra may suggest that they are displaying early signs of disease. Previous studies using SCP have shown that morphologically normal cells from patients with disease already show abnormal spectra patterns [24]. This is a testament to the sensitivity of SCP since it detects cells that have not yet displayed signs of morphological abnormalities as conventional cytopathology requires.

Although the majority of the cells appear morphologically normal, they share some compositional variations, which deviate from the biochemistry of normal cells. These compositional differences can be observed by analysis of the spectral data: we find that, in PCA, abnormal cells cluster together with cells with normal morphology from patients with disease. This is also supported by the ANN results which accurately classify morphologically normal cells with abnormal spectra into their correct pathological group.

Figure 2b depicts an overlay of the averaged second derivative, [vector](#) normalized spectra of the annotated cells. There is a significant trend in intensities of the amide I and phosphate bands: as disease progresses, the amide I decreases in intensity as the phosphate bands ( $\text{PO}_2^-$ ) generally increase. The decreased absorbance in the amide I manifold is most likely due to degradation in proteins and changes in the overall proteome of the cells which are shown to be reproducible throughout our spectra. In addition to protein degradation the formation of new proteins in the diseased samples is indicated by a shoulder apparent at  $1514\text{ cm}^{-1}$  in the amide II region.<sup>1</sup> The dysplastic cells show a considerable increase in the phosphate bands at  $1230$  and  $1080\text{ cm}^{-1}$  which may be indicative of increased rates proliferation and DNA replication.

Although there is some separation between the different classes discernible in Figure 2a the small size of the dataset does not permit a good visual discrimination. When using the entire dataset, as shown in Figure 3a, some separation is apparent but it is masked by the presence of a large num-

<sup>1</sup> It was pointed out by a reviewer that this frequency is generally associated with a tyrosine ring vibration. However, the intensity and width of the observed signal suggest that it is due to the amide II manifold.

ber of non-diagnostic squamous cells. This plot can be simplified substantially by including only cells that fell within a correlation coefficient of 0.99 of the mean annotated cell spectral classes. To achieve this, the mean spectra of the three annotated classes were computed, and only cells that were within 1% of the mean spectra were included in the PCA analysis shown in Figure 3b. Thus, Figures 3a and 3b represent the same PCA analysis, but in Figure 3b only the cells are included that fall within a narrow range of agreement with the annotated spectra shown in Figure 2a. Thus, Figure 3b may be viewed as a display of all cells in the dataset (annotated cells as well as non-annotated cells), but only cells that exhibit spectra very similar to the mean spectra of the annotated classes are shown, while less diagnostic cells were omitted from this plot. These 'non-diagnostic' cells will be discussed later in this paper.

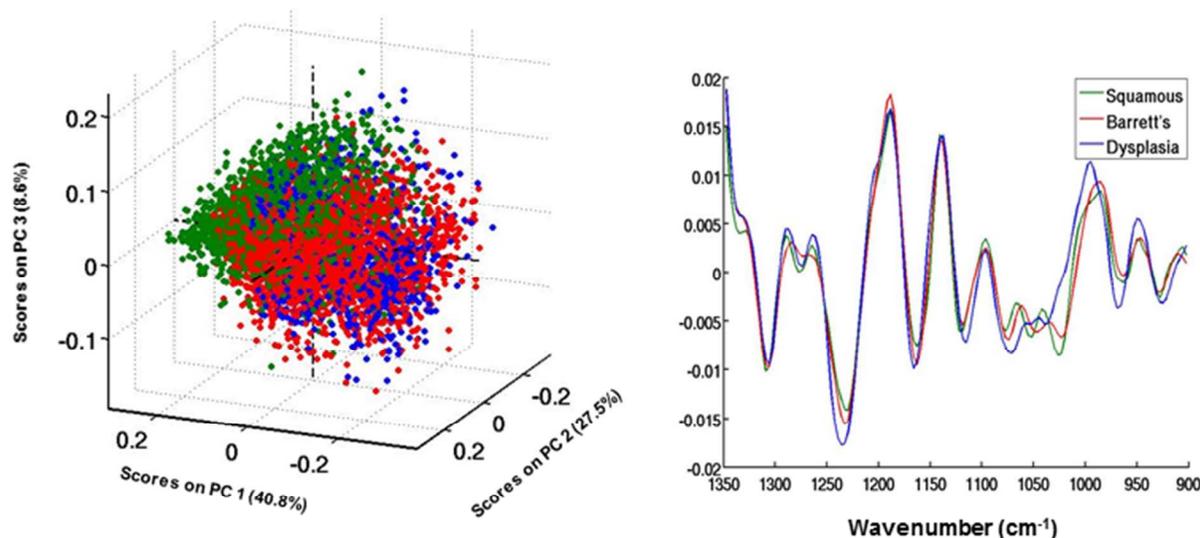
Figure 3a also shows that the squamous cells exhibit a diffuse cluster, indicating large variance in the cell spectra. This was also evident upon inspection of the spectra of all the squamous cells, shown in Figure 4a. This spectral plot indicates the presence of an amide I shoulder at  $1635\text{ cm}^{-1}$  that is generally associated with apoptosis or necrosis (see below). This shoulder is apparent in the spectra trace shown in Figure 3d. Thus, a spectral filter was written to eliminate all spectra with the feature at  $1635\text{ cm}^{-1}$ ; the filtered dataset is shown in Figure 4b, which shows vastly improved homogeneity. Thus, the diffuse cloud of squamous cells shown in Figure 3a is due to the presence of apoptotic or necrotic cells.

The signal at  $1635\text{ cm}^{-1}$  has prominently been observed for necrotic tissues [25], and may be due to the unfolding and precipitation of proteins. This signal was first reported

by Jamin, et al. [26] as a broad low frequency shoulder on the amide I peak. The second derivative spectra used in the present study displayed the 'necrosis' signal as a sharp shoulder ca.  $1635\text{ cm}^{-1}$  in the protein amide I manifold. The appearance of this peak is coupled with an increase of the peak at ca.  $1690\text{ cm}^{-1}$ , see Figure 4d. Whereas helical proteins exhibit a strong amide I peak at  $1650\text{--}1658\text{ cm}^{-1}$ , the combination of peaks at  $1635$  and  $1690\text{ cm}^{-1}$  is typically for  $\beta$ -sheet proteins. Upon precipitation of these protein sheets, the low component of the amide I manifold may shift to as low as  $1620\text{ cm}^{-1}$ [27].

We found that the cells displaying this 'necrotic' signal clustered together in PCA as shown in Figure 3c. We hypothesize that the necrotic signature is the result of proteins unfolding into  $\beta$ -sheets and precipitating. This signal was found to be so prominent in the spectra of some cells that classification into [their correct](#) disease classes was impossible. After removing the necrotic spectra the PCA 'scores plot' (Figure 5a) showed significant spectral differences among the cells from patients with healthy cells, Barrett's esophagus and dysplasia.

This separation between the classes indicates that there are identifiable biochemical differences between the disease states. Clearly these changes become more pronounced as the cells progress from healthy to Barrett's and to dysplasia. The biochemical variation occurring along PC1 is shown to be primarily due to changes in the amide I band, attributed to changes in protein composition (data not shown). This trend in the amide I may be due to degradation of proteins as cells respond to the disease. The major changes occur in the carbohydrate composition observed between  $1350$  and  $900\text{ cm}^{-1}$ (Figure 5b). The separation ob-



**Figure 5:** (a) PCA scores plot of data set shown in Figure 3a after removal of cells showing necrotic signatures. (b) Mean second derivative spectra of the three cell classes shown in Figure 5a.

served in these 'scores plots' may be partially due to the fact that in normal patients, a much higher percentage of squamous cells are harvested, and that the distinction of cell types is mostly due to the size difference between squamous and columnar cells. Cell-size based distinction of cells has been reported before by Romeo et al. [28], and it was argued that the spectral discrimination arose from different nucleus-to-cytoplasm ratios of small vs large cells; however this is not the case here as demonstrated by the similarity of the mean 2<sup>nd</sup> derivative spectra of the cell types. Furthermore the majority of cells in this data set are found to be morphologically normal and have approximately the same cell size.

Subtle changes in cellular biochemistry of the pathological classes [can be followed in the class mean spectra \(not shown\)](#). Each spectral trace is the mean of the second derivative, vector normalized spectra in each dataset. They represent 'snapshots' of the unique biochemistry of each pathological group. The Barrett's and dysplasia cells display slight intensity variations in both the amide I and amide II regions compared to cells exfoliated from squamous areas of the esophagus. The amide I band has a maximum at 1655  $\text{cm}^{-1}$ , and is somewhat decreased in the mean spectra of the diseased classes. The lower wavenumber shoulder of the amide II band is found at 1514  $\text{cm}^{-1}$  and is also shown to be slightly increased in both of the disease samples. The slight change in appearance of the amide I and II bands are most likely due to degradation of proteins and the expression of new proteins respectively. It is clear that as the disease progresses from Barrett's to dysplasia more pronounced changes in the protein region occur within the cells, which is supported by known biochemical evidence of protein conformational changes during carcinogenesis as well as changes in the overall proteome [29].

**Table II:** Binary classification sensitivity, specificity and overall accuracy of ANNs trained for supervised analysis..

	Sensitivity	Specificity	Accuracy
BE vs SQ	95.5%	94.7%	95.1%
BE vs DYS	88.7%	91.4%	90.0%
SQ vs DYS	93.4%	90.9%	92.1%

More evident variations occur in the 1350-900  $\text{cm}^{-1}$  region between all cell types and are emphasized by the low frequency region plot in Figure 5b. The blue spectrum shows slightly enhanced nucleic acid (DNA) features at ca. 1235 and 1080  $\text{cm}^{-1}$ ; furthermore, a peak at 1020  $\text{cm}^{-1}$  is missing in the blue spectrum. This latter peak is generally assigned to the C-O stretching and C-O-H deformation coordinates in carbohydrates, either in glycogen or glycoproteins. We observe a consistent decrease in the carbohy-

drate bands for Barrett's and dysplasia whereas we observe extremely strong absorption bands in the squamous cells; in fact the glycogen signal at 1020  $\text{cm}^{-1}$  is nearly absent in the dysplastic cells and decreased in the Barrett's cells. This decrease in glycogen tends to agree with known biochemical changes leading to tumor development such as increased energy consumption from cell division and has been reported previously for Barrett's associated changes by Stone, et al. [30, 31]. The diminished glycogen signal of cancerous and diseased cells has also been reported previously in the analysis of cervical cells [32].

Unsupervised analysis of the spectral data demonstrates that there are interpretable and reproducible changes that occur between the spectra of normal and diseased samples. In order to achieve a classification supervised, or trained algorithms, are used to predict class membership for individual cells. Here we used Artificial Neural Networks (ANN) to establish the sensitivity for this method when screening for esophageal disease.

The ANN was set up as follows: 400 randomly selected spectra from each class were assigned to a 'training set'. The remaining spectra in each class were subject to random selection and assigned to the 'testing set'. Spectra were selected so that equal numbers of spectra were present for each class in the 'testing set'. Training spectra were not used in testing. [However, due to the small number of patients in this study, the training and test spectra were not separated by patient; thus, cell spectra from the same patient were included in the training and test subsets.](#) Fifty (intensity) features were selected in each spectral vector via entropy-based feature selection, using 'rankfeature' function in MATLAB. The MATLAB 'feed-forward' ANN function was employed for ANN construction, utilizing 50 inputs, one hidden layer with two nodes and one output neuron (binary classification of normal/abnormal). The results of this ANN are expressed in terms of sensitivity, specificity and overall accuracy. Table II shows the results for the ANN classification.

The results from the ANN are a clear indication of the sensitivity and specificity that can be achieved with spectral methods. The overall accuracy is found to rival that of immunohistochemical staining, which is the most sophisticated and sensitive method conventional pathology has to offer; however this method still relies on the distribution of stains whereas spectral cytopathology does not. We wish to highlight that these results here were obtained from samples that contained mostly morphologically normal cells; that is, from cells that make up the vast majority in exfoliated samples. The observations here re-emphasize that SCP is able to detect biochemical changes associated with disease before morphological changes become apparent.

## DISCUSSION

Spectral Cytopathology (SCP) provides a non-destructive photonic approach, which takes a rapid measurement of a sample's biochemistry and reveals intrinsic biochemical changes that occur during the onset and progression of disease. SCP is a cytology based approach to identify disease in unstained cells using the principles of vibrational spectroscopy for analysis. An infrared spectrum from a single cell provides a biochemical snapshot, or immediate 'health status' of that particular cell. Molecular signatures, reflected in the infrared spectrum, are analyzed by unsupervised multivariate statistical methods, namely principal component analysis, to identify spectral patterns that exist between disease states. Biochemical components of a cell absorb infrared radiation at frequencies characteristic to particular bond types and functional groups present. Compounds such as proteins, nucleic acids and phospholipids have uniquely defined absorption bands within the 1800-900  $\text{cm}^{-1}$  range, and give what is known as a spectral 'fingerprint' for that given cell.

SCP has several advantages over conventional microscopy. Among them are its ability to provide an objective and reproducible diagnosis based on physical measurements of cellular biochemistry and, perhaps the most significant advantage, the ability to detect disease in cells before they have begun to display morphological changes. These 'morphologically normal' cells from patients with disease already show spectral abnormalities that can be related to the spectra of cells that were diagnosed as diseased. Thus, SCP is shown to identify spectral patterns that can aide in the earliest detection of cellular abnormalities.

This technique is ideal for diagnostics since it is a label-free method, requires minimal sample preparation, and is non-destructive, meaning that cells and tissues can be stained subsequent to infrared data acquisition, thus allowing spectral and classical pathology to be correlated.

In the author's laboratory, SCP has been used in several studies to establish the sensitivity of this method, such as screening for oral cancers, cervical and urinary and bladder cancers [13-15]. In all these studies, it was found that morphologically normal cells from patients with pre-cancerous disease exhibited spectra that resembled those of abnormal cells. Similar results had been obtained previously by Raman spectra methods [33]. Also, the spectral differences between cell types was established, for example, between squamous cells from the distal urethra and urothelial cells. These spectral differences were particularly pronounced in the second derivative spectra.

In this preliminary study we investigated SCP's potential as an enhancing screening method for esophageal disease. The aim of this study *was* to develop a methodology to be used in conjunction with conventional cytopathology that would assist cytopathologists in the diagnosis of esophageal

disease through collaborative work between clinicians and spectroscopists.

We report a *relatively* largest annotated data set of individual cellular spectra: 299 squamous, 145 Barrett's columnar and 14 dysplastic cells, totaling 458 diagnosed cells from 10 samples. The most abundant cell type found in each sample and disease class were squamous cells, which are more easily exfoliated than columnar cells.

Cells exfoliated from the esophagus of patients with healthy mucosa, Barrett's esophagus, and dysplasia were analyzed to establish the sensitivity of SCP when screening for esophageal disease. The biochemical signatures of disease are found to be reproducible throughout the majority of cells from each sample. Cells were subsequently stained and examined by an expert pathologist in order to build classification models to aid cytologists in rapid diagnosis of esophageal disease. We find that SCP of the esophagus works best when applied in conjunction with conventional pathology.

This study has demonstrated the ability of SCP to differentiate between esophageal cell types based on intrinsic molecular signatures. SCP provides a rapid measurement of cellular biochemistry and identifies reproducible spectral patterns that exist between disease states. Due to the inherent sensitivity towards changes in the biochemical composition of cells, SCP can provide additional unbiased diagnostic information to complement conventional pathology. These spectral changes can be used to distinguish between healthy and diseased specimens by analyzing large numbers of cells and training computer algorithms to differentiate between the two classes.

In this study we *took* a new approach in developing a synergistic methodology that would complement conventional cytopathology by *accurately* identifying diseased cells based on global changes in their biochemical composition. We report for the first time, findings, which indicate that SCP detects disease in cells from the esophagus that have no yet undergone any morphological changes associated with disease yet. The ability to reproducibly detect this spectral signature of disease in morphologically normal cells from abnormal samples adds a whole new dimension to the distinction of early stages of esophageal disease. We find that normal appearing cells from diseased samples still share the same biochemical characteristics as their abnormal counterparts. SCP was able to differentiate between all three of the pathological groups: healthy squamous, Barrett's oesphagus and dysplasia. Trends in the PCA 'scores' plots give clear indication on the progression of the disease from healthy to dysplastic and are reflected in the IR spectra. These spectral changes are found to be reproducible regardless of the morphological status of the cells. We find that the first changes that become apparent in the transition from healthy to Barrett's occur in the proteome of the cell. Further changes occur in the progression from Barrett's to

dysplastic as changes in DNA, RNA and carbohydrate composition. These biochemical changes agree with well known trends in tumor formation and carcinogenesis.

The ability of SCP to differentiate between disease states is evidence of the superior sensitivity of spectral methods over conventional pathology and can offer much

needed improvements in the screening process for esophageal disease.

### Acknowledgements

Partial support of this research from Grant CA 153148 from the National Institutes of Health (to MD) is gratefully acknowledged

### References

1. Spechler, S.J., *Barrett's Esophagus and Risk of Esophagus Cancer: A Clinical Review* JAMA, 2013. **310**(6): p. 627-636.
2. Souza, R.F. and S.J. Spechler, *Concepts in the Prevention of Adenocarcinoma of the Distal Esophagus and Proximal Stomach*. CA: A cancer journal for clinicians, , 2005. **55**(6): p. 334-351.
3. Barr, H., et al., *Endoscopic Screening and Surveillance for Barrett's Esophagus - Clinical Implications*. MedGenMed, 2006. **8**(2): p. 88-96.
4. Spechler, S.J., et al., *American Gastroenterological Association Medical Position Statement on the Management of Barrett's Esophagus*. Gastroenterology, 2011. **140**: p. 1084-1091.
5. Gerson, L.B., P.W. Groeneveld, and G. Triadafilopoulos, *Cost-effectiveness model of endoscopic screening and surveillance in patients with gastroesophageal reflux disease* Clin Gastroenterol Hepatol., 2004. **2**(10): p. 868-879.
6. Zhang, H.Y., S.J. Spechler, and R.F. Souza, *Esophageal Adenocarcinoma Arising in Barrett's Esophagus*. Cancer Letters, 2009. **275**: p. 170-177.
7. Wang, K.K. and R.E. Sampliner, *Diagnosis, Surveillance and Therapy of Barrett's Esophagus*. Am J Gastroenterol., 2008. **103**(103): p. 788-797.
8. Kara, M.A., et al., *High resolution endoscopy plus chromoendoscopy or narrow band imaging in Barrett's esophagus: a prospective randomized crossover study*. Endoscopy, 2005. **37**: p. 929-936.
9. Fitzgerald, R.C., et al., *British Society of gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus*. Gut, 2014. **63**(1): p. 7-42.
10. Montgomery, E., et al., *Reproducibility of the diagnosis of dysplasia in Barrett's oesophagus: A reaffirmation*. Hum Pathol., 2001. **32**: p. 368-378.
11. Bhat, S., *Barrett's oesophagus: risk of malignant progression and cost effectiveness of endoscopic surveillance*. 2012.
12. Kadri SR, et al., *Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care cohort study*. BMJ, 2010. **341**: p. c4372.
13. Papamarkakis, K., et al., *Cytopathology by Optical Methods: Spectral Cytopathology of the Oral Mucosa*. Laboratory Investigations, 2010. **90**: p. 589-598.
14. Bird, B., et al., *Cytology by Infrared Micro-Spectroscopy: Automatic Distinction of Cell Types in Urinary Cytology*. Vibr.Spectrosc., 2008. **48**(1): p. 101-106.
15. Schubert, J.M., et al., *Spectral Cytopathology of Cervical Samples: Detecting Cellular Abnormalities in Cytologically Normal Cells*. Laboratory Investigations, 2010. **90**: p. 1068-1077.
16. Bassan, P., et al., *The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications*. Analyst, 2013. **138**: p. 144-157.
17. Wrobel, T.P., et al., *Electric field standing wave effects in FT-IR transflection spectra of biological sections: Simulated models of experimental variability*. Vibr. Spectrosc., 2013. **69**: p. 84-92.
18. Miljković, M., et al., *Spectral Cytopathology: new aspects of data collection, manipulation and confounding effects*. Analyst, 2013. **138**: p. 3975-3982.
19. Schubert, J.M., et al., *Single Point vs. Mapping Approach for Spectral Cytopathology (SCP)*. J.Biophotonics, 2010. **3**(8-9): p. 588-596.
20. Bassan, P., et al., *Resonant Mie Scattering (RMieS) correction of infrared spectra from*

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- highly scattering biological samples. *Analyst*, 2010. **135**: p. 268-277.
21. IHCWorld. *Protocol OG6*. Available from: [http://www.ihcworld.com/protocols/special\\_stains/papanicolaou\\_stain.htm](http://www.ihcworld.com/protocols/special_stains/papanicolaou_stain.htm).
22. Lasch, P., et al., *Artificial Neural Networks as Supervised Techniques for FT-IR Microspectroscopic Imaging* *J.Chemometrics*, 2007. **20**(5): p. 209-220.
23. Adams, M.J., *Chemometrics in Analytical Spectroscopy*. 2nd ed. RSC Analytical Spectroscopy Monographs, ed. N.W. Barnett. 2004, Cambridge: Royal Society of Chemistry.
24. Cohenford, M. and B. Rigas, *Cytologically Normal Cells From Neoplastic Cervical Samples Display Extensive Structural Abnormalities on IR Spectroscopy: Implications for Tumor Biology*. *Proc. Natl. Acad. Sci.*, 1998. **95**: p. 15327-15332.
25. Akalin, A., et al., *Classification of Malignant and Benign Tumors of the Lung by Infrared Spectral Histopathology (SHP)*. *Laboratory Investigations*, 2015. **in press**.
26. Jamin, N., et al., *Chemical Heterogeneity in Cell Death: Combined Synchrotron IR and Fluorescence Microscopy Studies of Single Apoptotic and Necrotic Cells*. *Biopolymers (Biospectroscopy)*, 2003. **72** p. 366-373.
27. Kurouski, D., et al., *Is Supramolecular Filament Chirality the Underlying Cause of Major Morphology Differences in Amyloid Fibrils?* *J.Amer.Chem.Soc.*, 2014. **136**: p. 2302-2312.
28. Romeo, M.J., et al., *Infrared micro-spectroscopic studies of epithelial cells*. *Biochim Biophys Acta*, 2006. **1758**(7): p. 915-922.
29. Hanash, S. and A. Taguchi, *The grand challenge to decipher the cancer proteome*. *Nature Reviews Cancer*, 2010. **10**(9): p. 652-660.
30. Kendall, C., et al., *Evaluation of Raman probe for oesophageal cancer diagnostics*. *Analyst*, 2010. **135**: p. 3038-3041.
31. Stone, N., et al., *Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers*. *J.Raman Spectrosc.*, 2002. **33**(7): p. 564-573.
32. Chiriboga, L., Xie, P., Yee, H., Zarou, D., Zakim, D., Diem, M., *Infrared Spectroscopy of Human Tissue. IV. Detection of Dysplastic and Neoplastic Changes of Human Cervical Tissue Via Infrared Microspectroscopy*. *Cellular & Molecular Biology*, 1998. **44**(1): p. 219-229.
33. Utzinger, U., et al., *Near IR Raman Spectroscopy for in vivo detection of cervical precancers*. *Appl.Spectroscopy*, 2001. **55**(8): p. 955-959.