

# Evaluation of linear discriminant analysis for automated Raman histological mapping of esophageal high-grade dysplasia

## Joanne Hutchings

## Catherine Kendall

Gloucestershire Hospitals NHS Foundation Trust  
Leadon House  
Biophotonics Research Unit  
Great Western Road  
Gloucester, Gloucestershire, GL1 3NN,  
United Kingdom  
and  
Cranfield University  
Cranfield Health  
College Road  
Cranfield, Bedfordshire, MK43 0AL,  
United Kingdom

## Neil Shepherd

Cranfield University  
Cranfield Health  
College Road  
Cranfield, Bedfordshire, MK43 0AL,  
United Kingdom  
and  
Cheltenham General Hospital  
Gloucestershire Hospitals NHS Foundation Trust  
Gloucestershire Cellular Pathology Laboratory  
Sandford Road  
Cheltenham, Gloucestershire, GL53 7AN,  
United Kingdom

## Hugh Barr

Cranfield University  
Cranfield Health  
College Road  
Cranfield, Bedfordshire, MK43 0AL,  
United Kingdom  
and  
Gloucestershire Hospitals NHS Foundation Trust  
Department of Surgery  
Great Western Road  
Gloucester, Gloucestershire, GL1 3NN,  
United Kingdom

## Nicholas Stone

Gloucestershire Hospitals NHS Foundation Trust  
Leadon House  
Biophotonics Research Unit  
Great Western Road  
Gloucester, Gloucestershire, GL1 3NN,  
United Kingdom  
and  
Cranfield University  
Cranfield Health  
College Road  
Cranfield, Bedfordshire, MK43 0AL,  
United Kingdom

**Abstract.** Rapid Raman mapping has the potential to be used for automated histopathology diagnosis, providing an adjunct technique to histology diagnosis. The aim of this work is to evaluate the feasibility of automated and objective pathology classification of Raman maps using linear discriminant analysis. Raman maps of esophageal tissue sections are acquired. Principal component (PC)-fed linear discriminant analysis (LDA) is carried out using subsets of the Raman map data (6483 spectra). An overall (validated) training classification model performance of 97.7% (sensitivity 95.0 to 100% and specificity 98.6 to 100%) is obtained. The remainder of the map spectra (131,672 spectra) are projected onto the classification model resulting in Raman images, demonstrating good correlation with contiguous hematoxylin and eosin (HE) sections. Initial results suggest that LDA has the potential to automate pathology diagnosis of esophageal Raman images, but since the classification of test spectra is forced into existing training groups, further work is required to optimize the training model. A small pixel size is advantageous for developing the training datasets using mapping data, despite lengthy mapping times, due to additional morphological information gained, and could facilitate differentiation of further tissue groups, such as the basal cells/lamina propria, in the future, but larger pixels sizes (and faster mapping) may be more feasible for clinical application. © 2010 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3512244]

Keywords: esophageal cancer; high-grade dysplasia; diagnosis; Raman spectroscopy; imaging; mapping.

Paper 10266R received May 19, 2010; revised manuscript received Sep. 22, 2010; accepted for publication Sep. 28, 2010; published online Nov. 19, 2010.

## 1 Introduction

### 1.1 Clinical Motivation

The current gold standard for the diagnosis of esophageal precancers and cancers is endoscopic biopsy, followed by histological staining with hematoxylin and eosin (HE). Barrett's intestinal metaplasia (IM) is a cellular change induced in many individuals by acid reflux. Patients with Barrett's IM have an increased risk of developing esophageal adenocarcinoma. As a result, patients with Barrett's IM are often enrolled in endoscopic surveillance programs to detect precancers in the esophagus, which generate large numbers of tissue samples. This increases the workload for the histology department. In particular, large numbers of the samples taken are in fact classified as normal squamous or Barrett's IM, and therefore the diagnosis would not affect the way in which the patient is managed, since they would continue on the surveillance program. The current gold standard, histological diagnosis, is subjective and relies on the

---

Address all correspondence to: Nicholas Stone, Gloucestershire Hospitals NHS Foundation Trust, Leadon House, Biophotonics Research Unit, Great Western Road, Gloucester, Gloucestershire, GL1 3NN, United Kingdom. Tel: 44-8454-2254-86; Fax: 44-8454-2254-85. E-mail: n.stone@medical-research-centre.com.

interpretation of morphological features.<sup>1</sup> The majority of discrepancies occur in samples that are very similar, since there is a continuum of subtle cellular changes in the neoplastic progression to cancer, and consequentially, IM, low-grade dysplasia (LGD), and high-grade dysplasia (HGD) samples are of particular concern. A technique that can distinguish between these pathology groups would provide a clinically useful adjunct to current diagnostic methods.

Raman spectroscopy is an inelastic scattering technique that effectively provides a biochemical fingerprint, enabling the classification of different tissue types and pathology groups. Used in combination with multivariate analysis, the technique has the potential to provide automated, objective, and reproducible diagnosis of tissue pathologies. Our group has shown that the technique is a promising method to distinguish normal, pre-cancerous, and cancerous changes in unstained esophageal tissue using a laboratory-based Raman system.<sup>2,3</sup> Our group and others have demonstrated applications in other tissues, including the cervix,<sup>4</sup> larynx,<sup>5</sup> bronchus,<sup>6</sup> bladder,<sup>7,8</sup> colon,<sup>9</sup> breast,<sup>10</sup> skin,<sup>11</sup> and brain.<sup>12</sup> Further details can be found in recent review articles.<sup>13,14</sup> The technique also has the potential to be used *in vivo* through the use of fiber optic devices.<sup>15–18</sup>

## 1.2 Raman Spectral Mapping

It has previously been shown that Raman spectroscopy in combination with multivariate analysis can distinguish eight esophageal pathologies.<sup>2</sup> This work develops the idea of using the combination of multivariate analytical techniques and rapid Raman spectral mapping as a potential technique for automated histopathology. In previous publications, we have shown that technological advances have reduced Raman mapping times to a level that has made implementation in a clinical environment a future possibility.<sup>19,20</sup> Thus Raman spectroscopic mapping could potentially be used as an aid to the histopathologist. However, evaluation of linear discriminant analysis for automated analysis of such large datasets has yet to be carried out for Raman imaging applications. Raman mapping has advantages over current histology diagnosis, since stains are not required, and consequently this reduces sample preparation. A further question that remained unanswered was whether or not the possibility of additional information gained from high (lateral) spatial resolution Raman mapping would be a useful adjunct for the histopathologist. The issue of spatial resolution is discussed in this work.

Raman spectroscopy has the potential for high lateral spatial resolution mapping (micrometer to submicron level).<sup>21</sup> However, applications on biological tissue sections have been limited due to lengthy overall mapping times, often many hours for samples sized 1 mm<sup>2</sup>.

Fourier transform infrared (FT-IR) imaging is an alternative (vibrational spectroscopic) technique that provides biochemical information of biological tissue, as demonstrated by many studies.<sup>22–25</sup> Others have used both Raman and FT-IR as complimentary techniques.<sup>26,27</sup> The spatial resolution of laboratory-based spectrometers are diffraction limited: however, due to the wavelengths used for NIR Raman spectroscopy, this is less of a limitation, making higher ( $\sim 1 \mu\text{m}$ ) lateral spatial resolutions feasible.<sup>28</sup> The spatial resolution of laboratory-based FT-IR systems, however, varies with wavelength ( $\sim 8 \mu\text{m}$  at  $4000 \text{ cm}^{-1}$ ).<sup>29</sup> The use of synchrotron sources

has reduced the spatial resolution<sup>30</sup> by enabling a smaller spot size, but clinical applications are restricted due to practicalities of widespread implementation. Higher lateral spatial resolutions can also be achieved with attenuated total reflection (ATR) imaging, whereby a high refractive index crystal probes a smaller spot size. However, contact is required between the ATR crystal and the sample over the whole field of interest.

Spatial resolution is one of the most critical measurement parameters in spectroscopic imaging<sup>29</sup>: however, an in-depth quantification of the lateral spatial resolution is beyond the scope of this work. The lateral spatial resolution is often limited by the lateral step size parameter, since even for a system capable of achieving high lateral spatial resolution, if a step size greater than the lateral spatial resolution is used, the pixel size becomes the limiting factor. The term pixel size is therefore used in place of lateral spatial resolution in this work. For further information, the reader is referred to the aforementioned publications.

In previous studies by the authors using rapid Raman mapping,<sup>20</sup> the pixel size was not reduced below  $7.4 \mu\text{m}$  (approximately the width of the focused laser line using the  $50\times$  objective), due to limitations caused by the dataset size and also to prevent oversampling. Improvements in programming, software, and computer power have since enabled larger datasets to be handled. From one perspective, this has increased the total area that can be mapped, but from another perspective, this has also increased the spatial resolution that could be utilized. The detrimental effects of undersampling are clear with the possibility of missing an area of focal disease. This work explores Raman maps acquired with small pixel sizes and demonstrates the advantages for clinical diagnosis and automated classification. Oversampling is less significant than undersampling and can in fact be advantageous for spectroscopic mapping due to improvements in the definition of tissue boundaries.<sup>25</sup> The limitation of the system used in this study is a  $1.1\text{-}\mu\text{m}$  step size using a  $50\times$  objective, and since the advantage of synchronous readout technology, such as Renishaw's StreamLine<sup>TM</sup> (outlined elsewhere<sup>20</sup>) is greater for smaller step sizes, large datasets (of the order of hundreds of thousands of spectra) can be generated in a practicable time frame.

Many studies have reported the use of principal component analysis (PCA) for Raman imaging purposes.<sup>3,27,31,32</sup> To the best of the authors' knowledge, there are only two Raman mapping studies of esophageal tissue sections.<sup>3,20</sup> Shetty et al.<sup>3</sup> demonstrated high levels of glycogen in normal squamous tissue compared with a relative increase in the DNA levels in abnormal esophageal tissue sections. However, the Raman maps were acquired with large pixel sizes (potentially losing key spectral information) to enable sample coverage while minimizing long overall mapping times. Raman images presented in this study are of high spatial resolution and overcome previous limitations.

This work also extends the use of multivariate analysis to include PCA-fed linear discriminant analysis (LDA) of Raman images of esophageal tissue sections. Although PCA-fed LDA for pathology classification of Raman spectral data is a widely accepted technique, there are a limited number of publications applying the technique to Raman mapping of human tissue,<sup>11,33</sup> and only a few FT-IR imaging studies of biological tissue.<sup>34–37</sup> Other multivariate techniques such as cluster analysis (CA) and artificial neural networks (ANN) have been used for Raman and FT-IR image analysis. However, CA cannot be used

explicitly to generate diagnostic algorithms, because you cannot easily project independent data onto the CA model and predict which group the data belongs to. ANN have high computational requirements,<sup>38</sup> and there is a lack of transparency of the variables utilized by the discriminating algorithms. PCA-fed LDA has the advantage that it is well understood and enables the independent test dataset to be projected onto the training dataset. PCA/LDA also enables supporting biochemical information to be extracted from the PCA/LDA loadings, unlike “black box” methods that do not allow the user to validate the biochemical basis for separation. This work aims to evaluate the performance of a PCA-fed LDA model for automated histology classification, and comparison of these other analysis techniques with LDA is beyond the scope of this work.

## 2 Materials and Methods

### 2.1 Sample Collection and Preparation

Two samples from two different patients have been mapped in this study to evaluate the feasibility of the technique. Informed consent was obtained from patients undergoing routine upper gastrointestinal endoscopy and surgical resection. The Gloucestershire Local Research Ethics Committee granted ethical approval for this study.

Fresh tissue samples were obtained from endoscopic biopsy procedures and immediately snap frozen in liquid nitrogen. Biopsy samples are typically 1 to 2 mm in diameter. Samples were stored in a  $-80^{\circ}\text{C}$  freezer until measurements were carried out. For each sample, a  $15\text{-}\mu\text{m}$  frozen section was cut onto a (ultraviolet-grade) calcium fluoride ( $\text{CaF}_2$ ) substrate (Crystran, Poole, United Kingdom) for Raman spectral mapping. The thickness of the mapping section was chosen to maximize Raman scattered photons from the tissue section (while not taking the section beyond 1 to 2 cells thick).

A contiguous  $7\text{-}\mu\text{m}$  frozen section was cut and stained with HE for diagnosis by an expert gastrointestinal registry pathologist. The mapped section was also stained with HE following Raman spectral mapping.

The histological diagnosis was made using the contiguous HE section and subsequently verified on the HE stained mapped section (on  $\text{CaF}_2$ ) by a second histopathologist. Both the contiguous section HE and the mapped section (on  $\text{CaF}_2$  stained with HE) are shown for comparison. Regions of fibrous connective tissue (FCT), normal squamous (NSq), and high-grade dysplasia (HGD) were identified.

### 2.2 Raman Spectral Measurement

Raman maps were acquired using a customized Renishaw Raman System 1000 spectrometer with StreamLine™ technology (Renishaw Plc., Wotton-under-Edge, Gloucestershire, United Kingdom). The customized Raman system comprises a near-IR diode laser (830 nm,  $\sim 35\text{ mW}$  at the sample) for excitation, a Leica microscope with a Leica  $50\times$  (NA 0.5) long working distance objective to illuminate the sample (line focused using a fixed cylindrical lens) and collect the Raman scattered photons, a metal oxide edge filter to remove the elastically scattered light, a 300 lines/mm grating to disperse the inelastically scattered light, and a deep depletion charged-coupled device (CCD) detector. The StreamLine technology has been described in detail previously,<sup>20</sup> but in brief, the rapid mapping system utilizes

synchronous raster scanning of the sample across a line-focused laser spot ( $\sim 7\times 50\text{ }\mu\text{m}$ ) with CCD readout to allow faster spectral acquisition. The synchronous readout of the CCD is used to spatially separate spectral information acquired from different portions of the line focused laser, thus enabling the potential for sampling areas smaller than the length of the laser line. The pixel size (user defined) is determined by binning CCD pixels, therefore a pixel size (when back-projected through the microscope and spectrometer optics onto the sample) of any dimension can be chosen, limited only by the relative size of the CCD pixels and the magnification of the optics in the system. Pixels are usually defined as square pixels for convenience, but this is not essential. Selecting a pixel size less than the width of the laser line will result in oversampling, but this approach has been shown to improve spatial resolution.

A white light montage image of the tissue section on  $\text{CaF}_2$  was obtained using a  $2.5\times$  objective. The white light image was compared to the contiguous HE section to ensure the Raman map covered regions of different tissue type. Raman maps were then acquired ( $50\times$  objective) with step sizes of  $8.4$  and  $2.1\text{ }\mu\text{m}$  and an acquisition time of 15 s (to achieve spectra with good signal-to-noise ratio). This resulted in Raman maps with pixel sizes of  $8.4$  and  $2.1\text{ }\mu\text{m}$ . Overall mapping times depended on the area of the sample mapped, but were of the order of 2 to 4 h for  $8.4\text{-}\mu\text{m}$  pixel maps, and 12 to 18 h for  $2.1\text{-}\mu\text{m}$  pixel maps. Samples were air-dried prior to measurement.

### 2.3 Data Analysis

Saturated spectra were removed and cosmic rays were corrected by linear interpolation of the data points on either side of the cosmic ray peak. Subsequently, each map dataset was normalized and mean-centered. Principal component analysis (PCA) was carried out in Matlab (The MathWorks, Natick, Massachusetts) using the PLS toolbox (Eigenvector Technologies, Manson, Washington). Any remaining cosmic rays still evident in the PC loads and pseudocolor PC score images were blanked out, removed from the calculation, and the PCs regenerated.

### 2.4 Principal Component Imaging

Pseudocolor PC score maps were then plotted and overlaid to identify spectrally different regions within the Raman map. Each pixel of the PC scores image was color coded; the upper and lower extremes of the PC scores (0 to 25% and 75 to 100%) represented the pixels/spectra with the most significant contributions from the positive and negative aspects of the PC loads, respectively, and were illustrated with contrasting colors. Pixels falling into the central range of the scores were left transparent to enable the images to be overlaid. The corresponding PC loads were color coded accordingly to enable correlation of biochemical constituents from peaks within the PC loads with morphological information from the pseudocolor PC score image.

### 2.5 Comparison of Maps Acquired with Different Pixel Size

As a first step toward automated histopathology, bulk tissue discrimination was tested, i.e., discrimination between different

tissue types. Raman maps of sample 1 (which contained HGD and FCT) were acquired with two different pixel sizes (8.4 and 2.1  $\mu\text{m}$ ) to enable comparison. PC-fed LDA was then carried out (using the first ten PCs). Ten PCs were chosen as a cut-off, since beyond this the PC loadings represented only noise, thus minimizing the risk of the model fitting noise to the data. Based on the information from the histopathologist, spectra were classified as either calcium fluoride ( $\text{CaF}_2$ ), tissue border (TB), high-grade dysplasia (HGD) or fibrous connective tissue (FCT). Fluorescence (FI) was also included in the training model, since the PC imaging highlighted the presence of fluorescence spectra within the tissue structure. Since the fluorescence spectra are spectrally very different from Raman spectra, these were included as a separate group. The remaining spectra, for which their grouping was ambiguous (either due to the fact that there is not a distinct boundary between the tissue types, or the spectra were found to have overlapping PC load contributions), were excluded from the training dataset and labeled as the test dataset. The training dataset was normalized and mean-centred. Spectra acquired from  $\text{CaF}_2$  were included in the mean centering process, since it was concluded that the substrate would be an important contributor to the signals measured. Discrepancy with substrate impurities can lead to misclassifications, and furthermore, this can also be important for regions of thin tissue that may contain contributions from both substrate and tissue.

The test dataset was then scaled by subtracting the mean of the training dataset, and subsequently projected onto the LDA classification model as an independent test dataset. Although the test dataset is not truly independent, since it originates from the same sample, it is not included in the training dataset and therefore provides an adequate method of validating the model for this feasibility study.

Each pixel was then color coded according to the pathology group into which the classification model assigned each spectrum. The resulting LDA pseudocolor pathology map was then compared to the HE stained sections. Misclassified spectra were identified as black pixels.

### 2.6 Linear Discriminant Analysis Images Acquired Using Small Well-Defined Pathology Regions

To further test the limitations of the technique, the size of the training dataset was reduced such that only small, well-defined histological regions were selected. Since the majority of misclassifications in the previous model were in the TB group, this group was excluded from subsequent classification models.

A PCA-fed LDA model was generated using the training dataset, and the remainder of the map was assigned to the test dataset. PCA was carried out prior to LDA to reduce the number of variables and reduce noise. As described previously, the test dataset was scaled and projected onto the classification model, and an LDA pseudocolor pathology map was created.

### 2.7 Combining Maps with Different Pathology

Sample 2 was selected, as it contained NSq epithelium and FCT; a PCA-fed LDA pseudocolor pathology map was cre-

ated. As an initial method of evaluating the process, the maps from the two samples were combined to demonstrate the feasibility of the classification model when working for multiple samples. Regions were selected in the combined map to create a training dataset for the PCA-fed LDA model. The training dataset included two different  $\text{CaF}_2$  substrates, FCT from two different samples, HGD from one sample, and NSq from one sample.

### 2.8 Model Validation

The final six-group classification model was validated by randomly removing one third of the spectra in the training dataset and using this to test the model. This was repeated 200 times. The validated model accuracy was calculated by taking the average overall training performance for the 200 iterations. The sensitivity and specificity of each iteration was calculated and the mean used as the final validated sensitivity and specificity.

## 3 Results

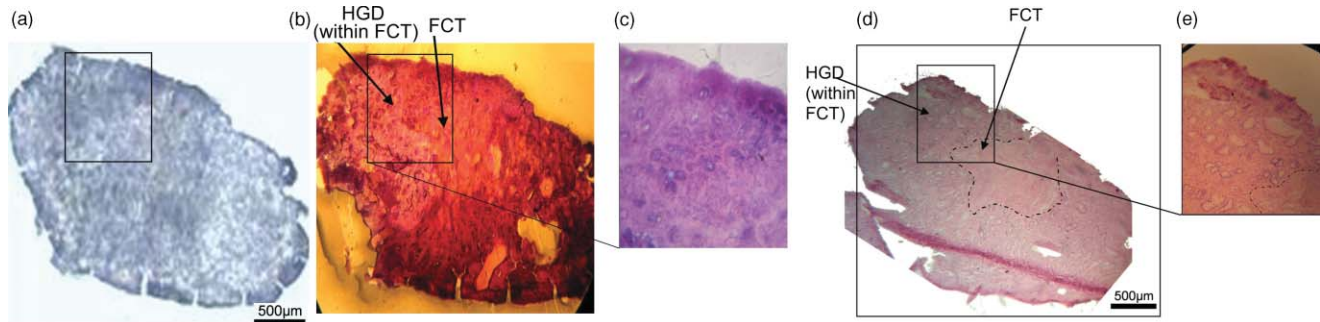
### 3.1 Principal Component Imaging

Figure 1 shows the white light montage image of sample 1, acquired using a 2.5 $\times$  objective. The box indicates the region containing HGD and FCT, which was mapped with two different pixel sizes. The histopathologist noted that the regions between the HGD glands were also FCT (interglandular FCT). The contiguous section and the mapped tissue section (on  $\text{CaF}_2$ ) stained with HE for histology purposes are also shown. The quality of the staining on the section that had been mapped previously was poorer than that of the contiguous section, as evident in Figs. 1(c), 1(d), and 1(e). However, the mapped section that was stained enabled better correlation of the morphological features of the tissue samples with those visible in the Raman map.

Figure 2 shows an example of a pseudocolor PC score image (PC 2) and corresponding PC load. The extremes (0 to 25% and 75 to 100%) of the color bar are represented by a single block of color, with the central portion remaining transparent. This enabled the different PC images to be overlaid, thus identifying morphological regions that could be assigned a pathology diagnosis by the histopathologist.

### 3.2 Comparison of Maps Acquired with Different Pixel Size

When the PC scores and loads were compared, both were found to be similar for the repeated maps. This suggests that there are no obvious biochemical changes occurring with time for the repeated maps; however, more subtle changes cannot be ruled out. The performance of the classification models was similar for maps acquired with both 8.4- and 2.1- $\mu\text{m}$  pixel size on sample 1. Overall training performances of 94.4% (79.4 to 99.0% sensitivity and 95.0 to 99.8% specificity) and 93.7% (87.2 to 100.0% sensitivity and 95.2 to 100.0% specificity) were obtained for the 8.4- and 2.1- $\mu\text{m}$  pixel size maps, respectively. The number of spectra correctly classified in each model is shown in Table 1. The remaining pixels (test dataset), indicated by white pixels, were projected onto the LDA model and then color coded



**Fig. 1** Sample 1. (a) White light image and mapped area. (b) Mapped tissue section (on  $\text{CaF}_2$ ) stained with HE. (c) Magnified region of mapped area. (d) Contiguous  $7\text{-}\mu\text{m}$  HE stained section. (e) Magnified region of (approximate) mapped area.

according to the LDA prediction. The LDA pseudocolor pathology maps are shown in Fig. 3 for both the  $8.4\text{-}\mu\text{m}$  and  $2.1\text{-}\mu\text{m}$  maps.

The LDA pseudocolor pathology maps are comparable for both pixel sizes when considering the bulk discrimination of HGD and FCT. There is a slight discontinuity in the  $2.1\text{-}\mu\text{m}$  pixel map at  $x$  pixel number 210, where two smaller maps were joined together. The spectral predictions in both are consistent with tissue pathology and location in the HE stained section image. The  $2.1\text{-}\mu\text{m}$  pixel map, however, demonstrates clearly defined boundaries of the HGD glands and gland lumen. Furthermore, it is evident the FCT can be seen to extend between the glandular features of the HGD, as confirmed by the histopathologist. In the  $8.4\text{-}\mu\text{m}$  pixel map, the HGD glands appear to be blurred with the surrounding FCT, and only small regions of interglandular FCT are identified.

The  $2.1\text{-}\mu\text{m}$  pixel map provides additional information relating to the sample morphology, which can be useful for identifying small features. In this example, additional information relating to structure of the sample can be gleaned in comparison with the  $8.4\text{-}\mu\text{m}$  pixel map. This indicates a loss of spectral information with decreased spatial resolution, which may require further work to confirm whether this is clinically significant or not. Additional work is also required to investigate the origins of the fluorescence within the maps, which appears to be structurally situated within the FCT.

Lasch and Naumann<sup>29</sup> rigorously investigated the effect of pixel size by binning adjacent pixels and comparing the resultant cluster images, but the different approach adopted in this study

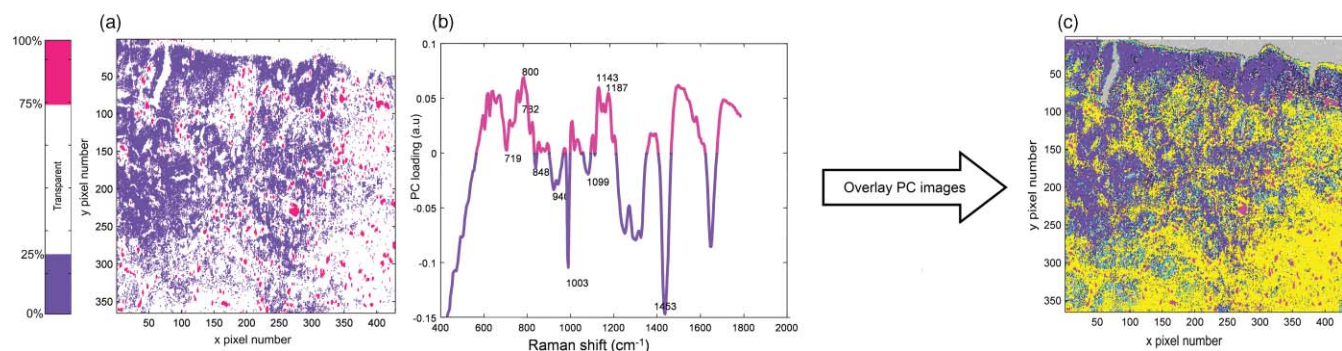
uses real experimental results, and therefore may be subject to any artifacts inherent within the CCD pixel binning. Since this is how the system would be used in real life, it was concluded that this was a more realistic approach for evaluating the system for clinical use.

### 3.3 Example of Normal Squamous Epithelium

The process outlined before was repeated on a map ( $15\text{-s}$  acquisition and  $8.4\text{-}\mu\text{m}$  pixel size) of sample 2 (containing NSq and FCT), but in this case the size of the training dataset was reduced further still, such that only small rectangular regions were selected.

Figure 4(a) shows the regions of the map selected for the training dataset (defined by small but distinct regions of NSq, FCT, and  $\text{CaF}_2$ ) color coded according to pathology. The spectra that were retained for the independent test dataset are represented as white pixels.

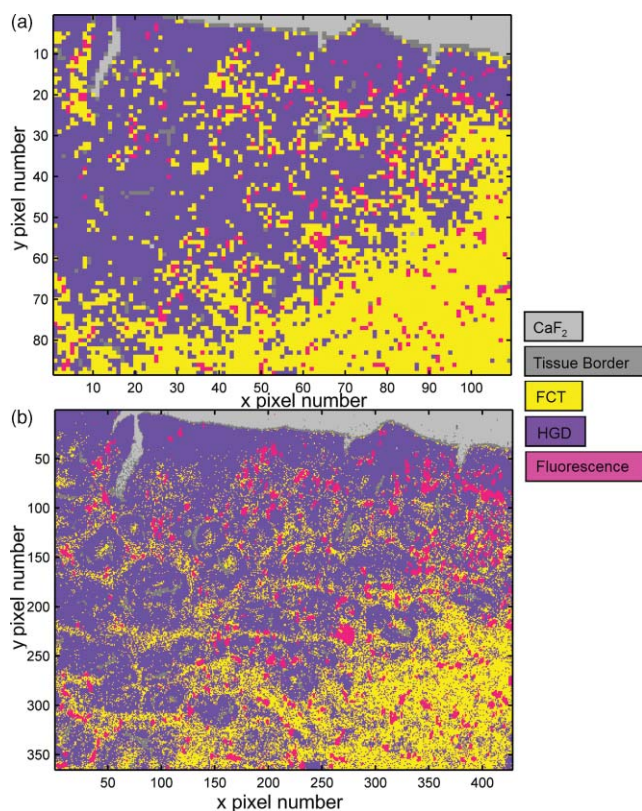
The overall training classification performance for the PCA-fed LDA model was 100%, as illustrated in Fig. 4(a), since there are no black pixels representing misclassified spectra. Fluorescence was not included as a group within the model, as this was not evident in the map. The projected model is shown in Fig. 4(b), superimposed as an inset on the white light image as a pseudocolor LDA image. The predicted pathology classification is represented by the color of the pixels. The HE images for the contiguous section and for the mapped section are also shown to illustrate the pathology of the mapped region.



**Fig. 2** Example of how the transparent PC images are generated and overlaid. (a) Pseudocolor PC 2 score image and (b) corresponding PC load for the  $2.1\text{-}\mu\text{m}$  map obtained from sample 1. (c) Overlaid pseudocolor PC score images (PCs 1 through 5).

**Table 1** Classification performance of the training dataset of the PCA-fed LDA model (8.4- and 2.1- $\mu\text{m}$  pixel sizes, 15-s acquisition time) measured on sample 1.

		CaF <sub>2</sub>	TB	FCT	HGD	Fl	Correctly classified (%)
8.4 $\mu\text{m}$	CaF <sub>2</sub>	374	7	0	0	0	98.2
	TB	14	150	24	1	0	79.4
	FCT	0	0	990	4	6	99.0
	HGD	0	3	196	3420	43	93.4
	Fl	0	2	0	0	165	98.8
2.1 $\mu\text{m}$	CaF <sub>2</sub>	5541	543	0	0	0	91.1
	TB	187	1352	11	0	0	87.2
	FCT	0	1	5399	0	0	100.0
	HGD	0	125	3951	65,813	714	93.2
	Fl	0	4	2	0	4067	99.9

**Fig. 3** Performance of the LDA tissue classification model (generated using sample 1: (a) 8.4- $\mu\text{m}$  pixel size and (b) 2.1- $\mu\text{m}$  pixel size). There is a subtle discontinuity (at x pixel number 210) in the 2.1- $\mu\text{m}$  map, since the large map area was generated by joining two smaller maps together.

### 3.4 Combining Maps with Different Pathology (2.1 $\mu\text{m}$ )

To further test the LDA projection of map data onto the tissue classification model, maps of the two samples (15-s acquisition time and 2.1- $\mu\text{m}$  pixel size) were combined to form a large map containing HGD (sample 1), FCT (from samples 1 and 2), NSq (sample 2), and CaF<sub>2</sub> (from samples 1 and 2). The map of sample 2 was cropped for convenience, since it allowed the dimensions of the two maps to be matched, illustrated by a solid black line across the map of sample 2. This would not be required for future applications. The cropped region was chosen to ensure that each classification group was represented in the map of sample 2.

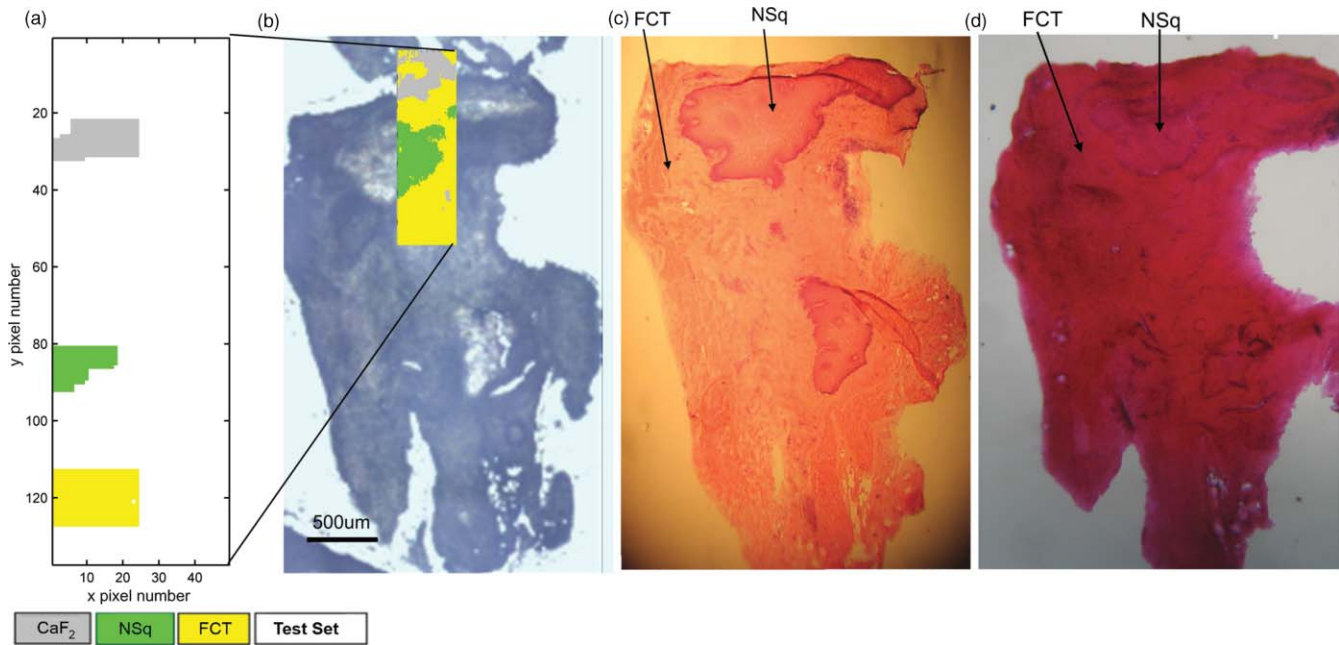
The combined map of sample 1 and 2 was then reanalyzed to investigate the feasibility of extending this to multiple tissue maps and tissue types. Again, small distinct regions of each tissue type (NSq, FCT, and HGD) and also CaF<sub>2</sub> and fluorescence were defined as the training dataset. The remainder of the dataset was retained as an independent test dataset, which was subsequently projected onto the classification model.

An initial training dataset (7640 spectra) model was used to separate five groups: normal squamous epithelium (NSq), fibrous connective tissue (FCT), high-grade dysplasia (HGD), substrate (CaF<sub>2</sub>), and fluorescence (Fl). The overall classification accuracy of the PCA-fed LDA model was 98.5% (95.8 to 99.9% sensitivity and 99.1 to 100% specificity). Projecting the test dataset (130,695 spectra) onto the classification model and reconstructing as a pseudocolor LDA image (Fig. 5) demonstrated good correlation with the HE stained sections; however, there are discrepancies that occur within the basal cells/lamina propria region of the NSq.

It is noteworthy that this incorrect classification only occurs when HGD is included in the training model. In Fig. 4(b), it is evident that the basal cells/lamina propria (BC/LP) are classified as NSq and FCT in the absence of HGD from the training model. This highlights a limitation in the way this LDA model was generated, since all the tissue groups available must be included as a group in the training dataset. Previous work published by Kendall et al.<sup>2</sup> demonstrated that eight and nine pathology groups can be distinguished with good sensitivity and specificity, but representing all other tissue types/pathologies remains a challenge. This also serves to highlight the strength of LDA in identifying spectra that relate to tissue regions which are biochemically similar, such as in this example, areas containing densely packed rapidly dividing cells.

Regions of fluorescence not previously identified in sample 2 are detected within the FCT. Again, the majority appear to be situated within the FCT tissue but further work is required to confirm this. Also, there are regions at the edge of sample 2 that are misclassified as HGD. It is thought this is due to tissue folds at the edge of the sample that may contain basal cells, although this is difficult to verify.

Conclusions can be drawn from the misclassification of the basal cells/lamina propria (BC/LP) as HGD, since the classification could be occurring based on biochemical signatures of cell nuclei that are rapidly proliferating and densely packed in both HGD and also the basal cells. Although potentially problematic if each of these groups are not included in future models, identification of the nuclear material in both the basal cells and



**Fig. 4** Pseudocolor LDA image of sample 2. (a) Regions selected from the mapped region for the training dataset, including normal squamous (green), fibrous connective tissue (yellow), and CaF<sub>2</sub> (gray). The white pixels represent the test dataset. (b) White light of the mapped sample with inset image representing the LDA projected classification for each pixel within the map. (c) HE image of 7-µm-thick consecutive section and (d) HE image of the 15-µm-thick mapped section (on CaF<sub>2</sub>). (Color online only.)

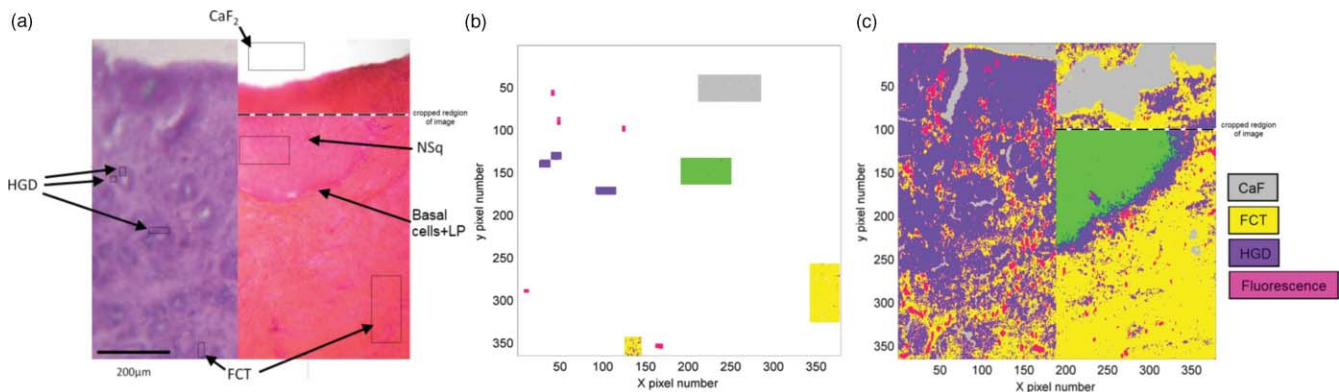
HGD does however demonstrate that the Raman spectroscopic technique is producing stain-free spectroscopic images equivalent to that produced by hematoxylin (from the HE histopathology stain). In the future, work is required to determine the optimum number of classification groups required for each application of spectral tissue diagnosis.

To further investigate the feasibility of using LDA for automated tissue classification, BC/LP was added as a separate spectral group within the model and the PC loads were analyzed to investigate the spectroscopic basis for tissue classification.

An initial training dataset (6483 spectra) validated model was used to separate six groups: normal squamous epithelium (NSq), fibrous connective tissue (FCT), high-grade dysplasia

(HGD), basal cells/lamina propria (BC/LP), substrate (CaF<sub>2</sub>), and fluorescence (Fl). HGD and BC/LP are easily separated using LDA with very few misclassifications (Table 2). The overall accuracy of model was 97.7% (sensitivity 95.0 to 100% and specificity 98.6 to 100%). Projecting the test dataset (131,672 spectra) onto the classification model and reconstructing as a pseudocolor LDA image [Figs. 6(a), 6(b), and 6(c)] demonstrated good correlation with the HE stained section, although there is a small region of FCT at the edge of the sample that is still misclassified as HGD.

To further investigate the biochemical basis of the classification, the PC scores for NSq, BC/LP, and HGD were plotted and the corresponding PC loadings analyzed for spectral features,



**Fig. 5** Pseudocolor LDA image of samples 1 and 2 (with central portion of map removed to match map dimensions). The dotted black line indicates the position where the sample 2 map was cropped. (a) HE of the contiguous section identifying regions selected for the training dataset. (b) The predicted classification for each pixel-normal squamous (green), fibrous connective tissue (yellow) and CaF<sub>2</sub> (grey), HGD (purple), fluorescence (Fl-pink). The white pixels (remainder of the map) represent the test dataset. Misclassified pixels/spectra are colour coded according to the LDA prediction. (c) Pseudocolor LDA image of sample 1 and sample 2 in which the test dataset is projected onto the LDA model.

**Table 2** Classification performance of the training dataset of the PCA-fed LDA model (2.1- $\mu\text{m}$  pixel sizes, 15-s acquisition time) measured on samples 1 and 2.

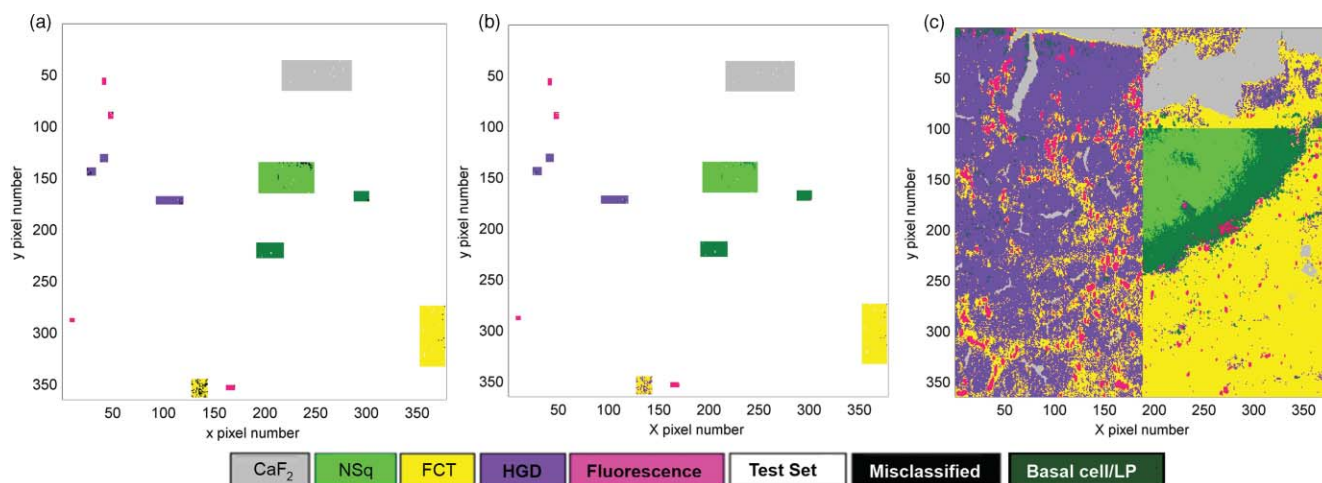
	NSq	HGD	CaF <sub>2</sub>	FCT	Fl	BC/LP	Total number of spectra (percent correctly classified)
NSq	1606	0	0	0	0	39	1645 (97.6%)
HGD	0	336	4	1	0	10	351 (95.7%)
CaF <sub>2</sub>	0	0	2063	0	0	0	2063 (100.0%)
FCT	0	82	0	1659	5	0	1746 (95.0%)
Fl	0	1	0	2	123	0	126 (97.6%)
BC/LP	1	2	0	1	0	548	552 (99.3%)

as shown in Fig. 7. It can be seen from Fig. 7(c) that PC1 accounts for the separation of BC/LP and HGD. Spectral Raman peaks can be identified in the PC1 loading at 852, 940, 1003, 1036, 1261, 1312, 1453, and 1659  $\text{cm}^{-1}$ . Tentative peak assignment can be made to collagen IV, which is a major biochemical constituent of the basement membrane. It is evident that PC2 separates BC/LP and HGD from NSq. The loading for PC2 exhibits one strong positive peak at 785  $\text{cm}^{-1}$  and a weaker peak at 1579  $\text{cm}^{-1}$ , which can tentatively be attributed to DNA, and multiple negative peaks (470, 855, 944, 1036, 1088, 1135, 1338, and 1467  $\text{cm}^{-1}$ ) demonstrate a strong correlation with glycogen peaks. It is expected that glycogen is present in NSq epithelium but not in abnormal tissue, since the cells are rapidly proliferating and use up the glycogen stores. This has been detected previously using Raman spectroscopy and reported previously in the literature.<sup>3</sup> An increase in DNA is also expected with HGD, since characteristically HGD contains enlarged densely packed nuclei. An increase in DNA can also be explained for the BC/LP, which contains densely packed and dividing cells close to the basement membrane.

#### 4 Discussion

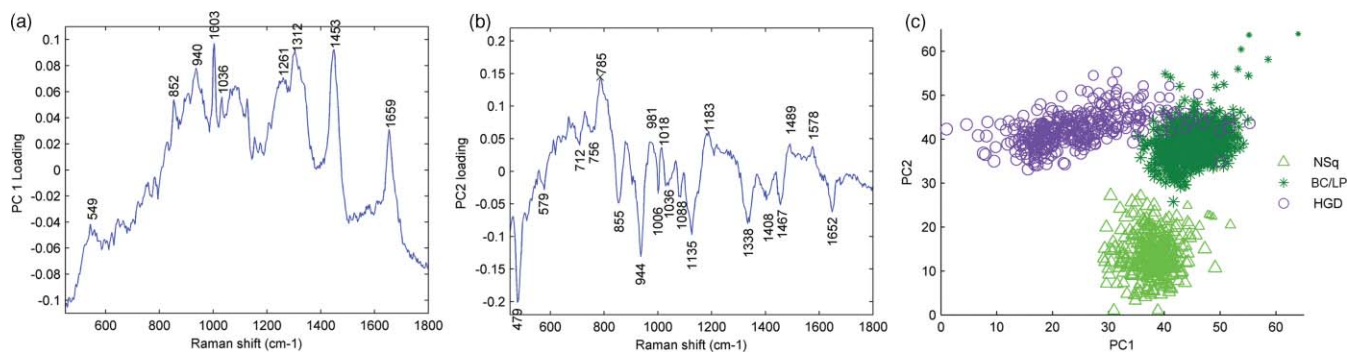
This study has shown that mapping with small pixel sizes (increased spatial resolution) is not necessarily needed for histological diagnosis, since bulk tissue pathology groups can be distinguished, even with a step size of 8.4- $\mu\text{m}$  pixel size. However, mapping with a small pixel size does have some advantages. There is the advantage of acquiring a large number of spectra, which is amplified for reduced step sizes due to the square relation between step size and number of pixels. The additional spatial and spectral biochemical information could potentially facilitate the separation of more pathology/tissue groups and potentially make classification models more robust, since spectral mixing between adjacent pixels is reduced.

There is also the potential, as discussed before, that the map obtained with high spatial resolution will identify more subtle biochemical features. As a result, the model performance for the small pixel size map could potentially be significantly better if the initial groupings are chosen more carefully. However, mapping at even smaller step sizes can also induce greater heterogeneity in the maps, even from cells of the same pathology



**Fig. 6** (a) Regions selected from the mapped region for the training dataset, including normal squamous (green), fibrous connective tissue (yellow), CaF<sub>2</sub> (gray), HGD (purple), basal cell/lamina propria (BC/LP, dark green), and fluorescence (Fl, pink). Black pixels identify those pixels misclassified by the LDA model. The white pixels (remainder of the map) represent the test dataset. (b) Same as (a), except the misclassified pixels/spectra are color coded according to the LDA prediction. (c) Pseudocolor LDA image of samples 1 and 2 (with central portion of map cropped to match map dimensions). (Color online only.)





**Fig. 7** (a) PC1 loading, (b) PC2 loading, and (c) PC scores scatter plot, PC1 versus PC2. HGD (purple circles), BC/LP (dark green stars), and NSq (light green triangles). (Color online only.)

(due to sampling different parts of the cell within each image pixel).

LDA is a well known and accepted technique for spectral classification, and this work has shown its potential application in Raman imaging for histological diagnosis. It has also served to highlight limitations of this study (in the sense that spectra, such as spectra from the BC/LP, may be misclassified if not included in the original training dataset). Further work is still required to investigate the extent to which the technique can be exploited with respect to automated imaging. Larger sample numbers and the full range of pathology/tissue groups (Barrett's IM, low-grade dysplasia, adenocarcinoma, etc.) need to be included in the training dataset. The projection of a test dataset onto the model allows it to be validated, but more rigorous validation and testing using further samples (and pathologies) will still be required. Only two samples from two patients were included in this study, but previous work has demonstrated that classification is feasible over a wider population with 1125 spectra (point measurements) on 87 homogeneous biopsy samples from 44 patients.<sup>2</sup> Nevertheless, this is an important step in the move toward clinical implementation of vibrational spectroscopy in combination with multivariate analysis for automated histopathology.

It is well known that the initial group choices are an important factor with LDA model performance. Therefore, this study was limited by operator choice in the selection of the training dataset. Cluster analysis may provide an alternative method, but since the technique is computationally intensive due to the large dataset sizes, this was not feasible in this case. Furthermore, cluster analysis also relies on prior knowledge of the number of groups present, and relies on the operator setting this parameter. Further work is also required to optimize selection of spectra for the classification model. Cross-sample test dataset validation demonstrated that this technique could be applied across different samples.

The degree of classification required for specific pathology applications remains an unanswered question. For example, if the ultimate aim is only to distinguish normal from abnormal tissue, then relatively crude spatial averaging and poor signal-to-noise spectra could be used. However, if the aim is to separate out more subtle changes such as tissue types, precancers, cancers, and even predict prognosis, then more subtle biochemical features may need to be resolved. It will also depend on whether the histopathologist is confident in spectroscopic diagnosis, without the additional morphological information repre-

sented in the form of a pseudocolor histology image. If not, then maps acquired with a crude pixel size could potentially provide a rapid and automated method of pathology diagnosis. A combination of modalities may be advantageous, for example, FT-IR for mapping the entire sample, followed by small pixel size Raman mapping of regions of interest. The complementary nature of FT-IR and Raman is being explored by many groups, including ours.<sup>26,27</sup>

The extent to which we attempt to separate out pathology information is a question of clinical need, which ultimately will need to be answered by the histopathologist.

## 5 Conclusions

This study shows that the LDA projection imaging process can potentially be applied to multiple mapped samples for automated histopathology, but further work is required before clinical implementation of the technique, and other analysis methods may be more suitable. Furthermore, spatial information can be obtained by visually representing LDA classification as pseudocolor images. Displaying LDA classification models in this way can provide insightful information that may help to explain misclassifications based on morphological features, which is not possible from a traditional scatter plot representation.

It also appears from this initial study that mapping with a small pixel size is not essential for clinical diagnosis of bulk tissue types, but has advantages for discriminating further tissue types and enabling better correlation with tissue morphology for classification model training. Further work is required to develop diagnostically relevant algorithms in close collaboration with clinicians.

## Acknowledgments

The authors would like to thank Jo Motte and Christine Braune for their assistance with sample preparation and Linmarie Ludeman for providing additional histology diagnosis. Financial support from the Institute of Physics and Engineering in Medicine (IPEM) Research Training Fellowship, Royal Society Dorothy Hodgkin Fellowship, and National Institute of Health Research (Post-Doctoral and Career Scientist Fellowships) is gratefully acknowledged.

## References

1. R. C. Haggitt, "Barrett's esophagus, dysplasia, and adenocarcinoma," *Hum. Pathol.* **25**(10), 982–993 (1994).

2. C. Kendall, N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett, and H. Barr, "Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's esophagus," *J. Pathol.* **200**, 602–609 (2003).
3. G. Shetty, C. Kendall, N. Shepherd, N. Stone, and H. Barr, "Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of esophagus," *Brit. J. Cancer* **94**, 1460–1464 (2006).
4. A. Robichaux-Viehoever, E. Kanter, H. Shappell, D. Billheimer, H. Jones, and A. Mahadevan-Jansen, "Characterization of Raman spectra measured *in vivo* for the detection of cervical dysplasia," *Appl. Spectrosc.* **61**(9), 986–993 (2007).
5. N. Stone, P. Stavroulaki, C. Kendall, M. Birchall, and H. Barr, "Raman spectroscopy for early detection of laryngeal malignancy: preliminary results," *Laryngoscope* **110**(10), 1756–1763 (2000).
6. S. Koljenović, T. C. B. Schut, J. P. van Meerbeeck, A. P. W. M. Maat, S. A. Burgers, P. E. Zondervan, J. M. Kros, and G. J. Puppels, "Raman microspectroscopic mapping studies of human bronchial tissue," *J. Biomed. Opt.* **9**(6), 1187–1197 (2004).
7. P. Crow, A. Molckovsky, N. Stone, J. Uff, B. Wilson, and L. M. WongKeeSong, "Assessment of fiberoptic near-infrared Raman spectroscopy for diagnosis of bladder and prostate cancer," *Urology* **65**(6), 1126–1130 (2005).
8. B. W. D. de Jong, T. C. B. Schut, K. Maquelin, T. van der Kwast, C. H. Bangma, D. J. Kok, and G. J. Puppels, "Discrimination between nontumor bladder tissue and tumor by Raman spectroscopy," *Anal. Chem.* **78**, 7761–7769 (2006).
9. J. C. Taylor, C. A. Kendall, N. Stone, and T. A. Cook, "Optical adjuncts for enhanced colonoscopic diagnosis," *Brit. J. Surg.* **94**(1), 6–16 (2007).
10. J. Kneipp, T. B. Schut, M. Kliffen, M. Menke-Pluijmers, and G. Puppels, "Characterization of breast duct epithelia: a Raman spectroscopic study," *Vib. Spectrosc.* **32**(1), 67–74 (2003).
11. M. Larraona-Puy, A. Ghita, A. Zoladek, W. Perkins, S. Varma, I. H. Leach, A. A. Koloydenko, H. Williams, and I. Notingher, "Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma," *J. Biomed. Opt.* **14**(5), 054031 (2009).
12. S. Koljenović, T. B. Schut, A. Vincent, J. M. Kros, and G. J. Puppels, "Detection of meningioma in dura mater by Raman spectroscopy," *Anal. Chem.* **77**(24), 7958–7965 (2005).
13. C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker, and N. Stone, "Vibrational spectroscopy: a clinical tool for cancer diagnostics," *Analyst* **134**(6), 1029–1045 (2009).
14. C. Krafft, G. Steiner, C. Beleites, and R. Salzer, "Disease recognition by infrared and Raman spectroscopy," *J. Biophoton.* **2**(5), 13–28 (2009).
15. Y. Komachi, H. Sato, Y. Matsuura, M. Miyagi, and H. Tashiro, "Raman probe using a single hollow waveguide," *Opt. Lett.* **30**(21), 2942–2944 (2005).
16. Z. Huang, M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, and K. G. Yeoh, "in vivo early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy," *J. Biomed. Opt.* **15**(3), 037017 (2009).
17. J. T. Motz, S. J. Gandhi, O. R. Scepanovic, A. S. Haka, J. R. Kramer, R. R. Dasari, and M. S. Feld, "Real-time Raman system for in vivo disease diagnosis," *J. Biomed. Opt.* **10**(3), 031113 (2005).
18. J. C. C. Day, R. Bennett, B. Smith, C. Kendall, J. Hutchings, G. M. Meaden, C. Born, S. Yu, and N. Stone, "A miniature confocal Raman probe for endoscopic use," *Phys. Med. Biol.* **54**(23), 7077–7087 (2009).
19. J. Hutchings, C. Kendall, N. Shepherd, B. Smith, H. Barr, and N. Stone, "Rapid Raman microscopic imaging for potential histological screening," *Proc. SPIE* **6853**, 685305 (2008).
20. J. Hutchings, C. Kendall, B. Smith, N. Shepherd, H. Barr, and N. Stone, "The potential for histological screening using a combination of rapid Raman mapping and principal component analysis," *J. Biophoton.* **2**(1–2), 91–103 (2009).
21. S. Schlücker, M. D. Schaeberle, S. W. Huffman, and I. W. Levin, "Raman microspectroscopy: a comparison of point, line, and wide-field imaging methodologies," *Anal. Chem.* **75**(16), 4312–4318 (2003).
22. P. Lasch, W. Haensch, D. Naumann, and M. Diem, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis," *BBA-Mol. Basis Dis.* **1688**(2), 176–186 (2004).
23. B. Bird, M. Miljkovic, M. J. Romeo, J. Smith, N. Stone, M. W. George, and M. Diem, "Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology," *BMC Clin. Pathol.* **8**, 8 (2008).
24. M. J. Baker, E. Gazi, M. D. Brown, J. H. Shanks, N. W. Clarke, and P. Gardner, "Investigating FTIR based histopathology for the diagnosis of prostate cancer," *J. Biophoton.* **2**(1–2), 104–113 (2009).
25. R. Bhargava, "Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology," *Anal. Bioanal. Chem.* **389**(4), 1155–1169 (2007).
26. C. Krafft, D. Codrich, G. Pelizzo, and V. Sergio, "Raman and FTIR imaging of lung tissue: Methodology for control samples," *Vib. Spectrosc.* **46**(2), 141–149 (2008).
27. M. Isabelle, N. Stone, H. Barr, M. Vipond, N. Shepherd, and K. Rogers, "Lymph node pathology using optical spectroscopy in cancer diagnostics," *Spectrosc.* **22**(2–3), 97–104 (2008).
28. *Infrared and Raman Spectroscopic Imaging*, R. Salzer and H. W. Siesler, Eds., Wiley-VHC, New York (2009).
29. P. Lasch and D. Naumann, "Spatial resolution in infrared microspectroscopic imaging of tissues," *BBA-Biomembranes* **1758**(7), 814–829 (2006).
30. L. M. Miller and P. Dumas, "Chemical imaging of biological tissue with synchrotron infrared light," *BBA-Biomembranes* **1758**(7), 846–857 (2006).
31. C. A. Hayden and M. D. Morris, "Effects of sampling parameters on principal components analysis of Raman line images," *Appl. Spectrosc.* **50**(6), 708–714 (1996).
32. C. Krafft, S. B. Sobottka, G. Schackert, and R. Salzer, "Raman and infrared spectroscopic mapping of human primary intracranial tumors: a comparative study," *J. Raman Spectrosc.* **37**(1–3), 367–375 (2006).
33. S. Koljenović, L. P. Choo-Smith, T. C. Bakker-Schut, J. M. Kros, H. J. van den Berge, and G. J. Puppels, "Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy," *Lab. Invest.* **82**(10), 1265–1277 (2002).
34. J. R. Mansfield, L. M. McIntosh, A. N. Crowson, H. H. Mantsch, and M. Jackson, "LDA-guided search engine for the nonsubjective analysis of infrared microscopic maps," *Appl. Spectrosc.* **53**(11), 1323–1330 (1999).
35. C. Krafft, S. B. Sobottka, K. D. Geiger, G. Schackert, and R. Salzer, "Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis," *Anal. Bioanal. Chem.* **387**(5), 1669–1677 (2007).
36. C. Beleites, G. Steiner, M. Sowa, R. Baumgartner, S. Sobottka, G. Schackert, and R. Salzer, "Classification of human gliomas by infrared imaging spectroscopy and chemometric image processing," *Vib. Spectrosc.* **38**(1–2), 143–149 (2005).
37. S. B. Sobottka, K. D. Geiger, R. Salzer, G. Schackert, and C. Krafft, "Suitability of infrared spectroscopic imaging as an intraoperative tool in cerebral glioma surgery," *Anal. Bioanal. Chem.* **393**(1), 187–195 (2009).
38. P. Lasch, M. Diem, W. Hänsch, and D. Naumann, "Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging," *J. Chemometr.* **20**(5), 209–220 (2006).