Journal of Gastroenterology

AUTOMATED CYTOLOGICAL DETECTION OF BARRETT'S NEOPLASIA WITH INFRARED SPECTROSCOPY

--Manuscript Draft--

Manuscript Number:	JOGA-D-16-00906R1
Full Title:	AUTOMATED CYTOLOGICAL DETECTION OF BARRETT'S NEOPLASIA WITH INFRARED SPECTROSCOPY
Article Type:	Original Article
Section/Category:	Experimental - Upper GI Tract
Manuscript Classifications:	10.010: Esophagus - Barrett's esophagus; 10.080: Esophagus - esophageal cancer
Funding Information:	Royal College of Surgeons of England (Surgical Research Fellowship) — Mr Oliver Jamieson Old
Abstract:	Background Development of a non-endoscopic test for Barrett's esophagus (BE) would revolutionize population screening and surveillance for patients with BE. Swallowed cell collection devices have recently been developed to obtain cytology from the esophagus: automated detection of neoplasia in such samples would enable large-scale screening and surveillance. Methods Fourier Transform infrared spectroscopy (FTIR) was used to develop an automated tool for detection of BE and Barrett's neoplasia in esophageal cell samples. Cytology brushings were collected at endoscopy, cytospun onto slides and FTIR images measured. An automated cell recognition program was developed to identify individual cells on the slide. Results Cytology review and contemporaneous histology was used to inform a training dataset containing 141 cells from 17 patients. A classification model was constructed using principal component analysis (PCA) fed linear discriminant analysis (LDA), then tested using leave one sample out cross validation (LOSOCV). Applying this training model to whole slide samples, a threshold voting system was used to classify samples according to their constituent cells. Across the entire dataset of 115 FTIR maps from 66 patients, whole samples were classified with sensitivity and specificity respectively as follows: normal squamous 79.0% and 81.1%, non-dysplastic Barrett's 31.3% and 100%, and neoplastic Barrett's 83.3% and 62.7%. Conclusions Analysis of esophageal cell samples can be performed using FTIR with reasonable sensitivity for Barrett's neoplasia, though poor specificity with the current technique.
Corresponding Author:	Oliver Jamieson Old, MD MBChB BMedSc Bristol Royal Infirmary Bristol, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Bristol Royal Infirmary
Corresponding Author's Secondary Institution:	
First Author:	Oliver Jamieson Old, MD MBChB BMedSc
First Author Secondary Information:	
Order of Authors:	Oliver Jamieson Old, MD MBChB BMedSc
	Gavin Lloyd, PhD
	Martin Isabelle, PhD
	Max Almond

	Catherine Kendall
	Karol Baxter
	Neil Shepherd
	Angela Shore
	Nick Stone
	Hugh Barr
Order of Authors Secondary Information:	
Author Comments:	Dear Editors, Many thanks for the opportunity to resubmit our article. We have addressed each of the concerns raised by the reviewers and we feel we have improved our manuscript in the process. Yours faithfully, Oliver Old Surgical Research Fellow Biophotonics Research Unit Gloucestershire Hospitals NHS Foundation Trust
Response to Reviewers:	Thank you for the opportunity to resubmit our article - please see attached file 'Response to Reviewers'.

Response to reviewers

Many thanks for your comments and the opportunity to resubmit our article. Our responses are highlighted in red below.

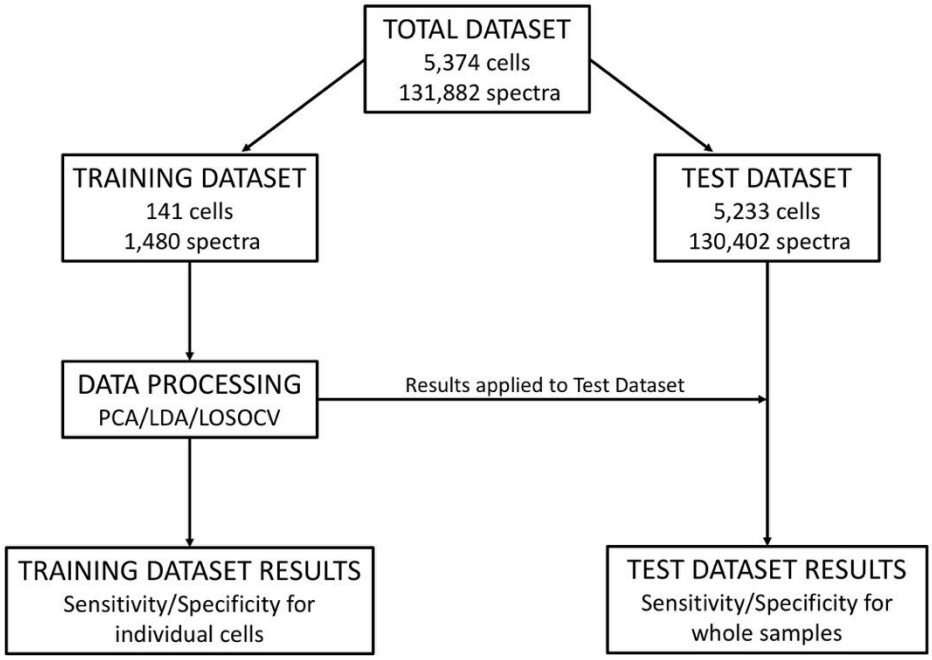
Reviewer #1: I have been asked to review the submitted manuscript entitled "Automated cytological detection of Barrett's neoplasia with infrared spectroscopy" by Dr. Old et al.. They evaluated the accuracy for cytological diagnosis of Barrett's esophagus, dysplasia, and adenocarcinoma using Fourier Transform infrared spectroscopy (FTIR). I agree with the clinical importance of automated diagnostic methods of Barrett's metaplasia and carcinoma. FTIR looks promising but further investigation seems to be needed. However, there are several concerns especially in the study design.

1. The training set have to be separated from the test set. I could not clearly figure out whether the 17 patients in training set were included in the whole 66 patients. If the training dataset were included in the test set, overfitting results could be shown. Flow diagram should be added to explain each cohort.

No spectra included in the training set were included in the test dataset, as stated in the text in the section headed 'Test dataset' in the Results:

'Cells that were included in the training dataset were excluded from the test set, to prevent overestimation of performance.'

I have moved this sentence to the Methods section for clarity. The following flow diagram (as suggested) could be added immediately after this sentence to aid understanding, however we have reached the limit of the number of figures allowed for this manuscript, so I cannot attach a further figure. This could be included as Figure 1 (with renaming of later figures) if the editors will allow.



2. Gold standard for this study should be the pathological diagnosis using biopsy samples. Then the diagnostic accuracy should be compared between FTIR and the classical cytological diagnosis.

The gold standard for comparison for the test set in this study was the biopsy result, as suggested. We have highlighted this with a further sentence in the 'Test dataset' section as follows:

'This predicted pathology was compared against the biopsy result from the same region and sensitivity and specificity calculated using this gold standard.'

Not every cell was examined cytologically (there were 5,374 cells in total) as this was felt to be impractical. This was however used for every cell included in the training dataset as outlined in the 'Training dataset' section:

'cells whose classification on appearance by two cytopathologists was in agreement with contemporaneous endoscopy and biopsy from the same region were included.'

3. There is no description that samples of low-grade dysplasia were included in which group.

We have added the following sentence in the 'Training dataset' section:

'The dysplasia/adenocarcinoma group included any degree of dysplasia (low or high grade), but samples classified as 'indefinite for dysplasia' were excluded.'

4. A scatter plot (like fig. 1) of the test set should be shown.

The analysis of individual cells in the test dataset is more complicated than the training dataset: cells in the training set have a known cytological appearance that matches the histology taken from the same locus. Since every cell has a known pathology label, a scatter plot can be produced using the known label for every cell. In the test dataset, individual cells do not have a known pathology label. The pathology of the whole slide is known, but the individual cells on the slide may be a heterogeneous population. For this reason we used a threshold voting system to assign predicted pathology to whole slide samples based on the number of cells per slide of a given predicted pathology. It is not possible to present this data with a scatter plot in the same way as Figure 1 of the training dataset.

5. It looks difficult to discriminate the three groups at once. Since the diagnostic accuracy for normal squamous cells seems to be comparably high, I recommend the diagnostic method should be separated into two steps: first, distinguish columnar cells from squamous cells; second, distinguish dysplasia/adenocarcinoma from metaplastic cells. Such physiologically reasonable strategy might elevate the accuracy. Even if you can achieve good performance only in the first step, this could be clinically useful for the screening of high risk group.

We attempted this in our initial analysis and it gave similar results to the 3 group model presented in our paper. The results from our initial 2 group model (differentiating normal squamous from columnar cells, as suggested by Reviewer 1) were as follows:

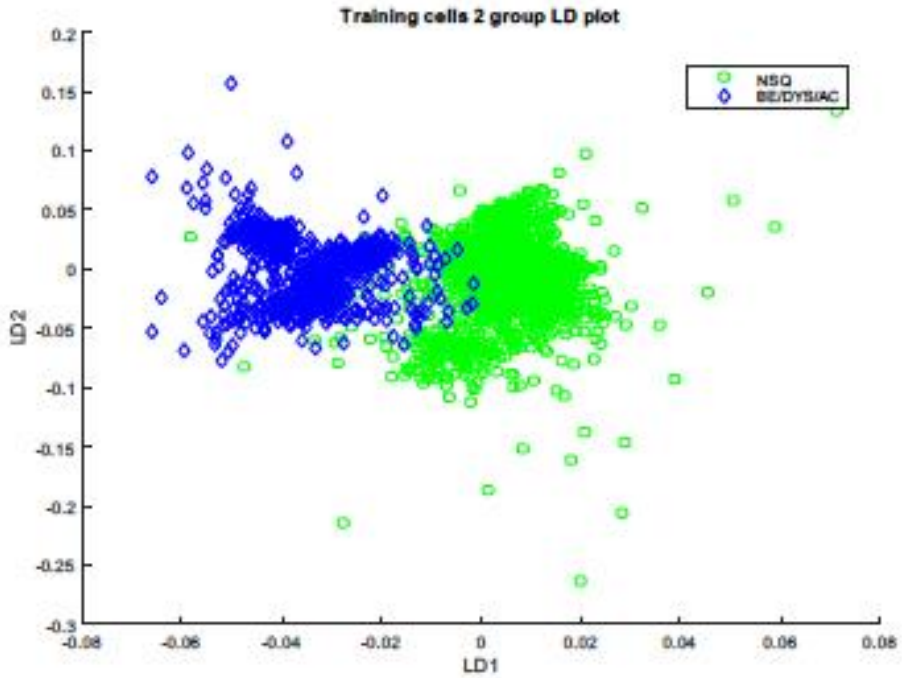


Figure 5-27 Scatter plot of all spectra, plotted by linear discriminant (LD) function. NSQ = normal squamous, BE = Barrett's oesophagus, DYS = dysplasia, AC = adenocarcinoma.

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

Mean values after 50 iterations		**NSQ**	**BE/DYS/AC**
		Mean % (SD)	**Mean % (SD)**
Individual spectra	Sensitivity	**84.4 (5.8)**	**72.6 (2.6)**
	Specificity	**72.6 (2.6)**	**84.4 (5.8)**

Since the values were similar to those from the 3 group model, and including a further table and graph lengthens the paper further, we decided not to present this additional data in our manuscript. We have added the following sentence to the discussion to indicate this:

'A two-stage analysis may be developed whereby cells are initially separated into squamous or columnar by a predictive model, and then a further analysis used to separate dysplastic

and non-dysplastic cells. A two-stage model was attempted using the current dataset, giving very similar results to the 3-group model presented here (two-stage results not shown).'

Reviewer #2: Summary
Old et al. developed automated cytological detection of esophageal adenocarcinoma or dysplasia in Barrett's esophagus. This paper is a well-written interesting study.

1. Although the authors discussed, reasons why the authors set small number of the training dataset of Barrett's esophagus are not clear.

We have elaborated on this further in the Discussion section as follows:

'The major limiting factor in this study was the small number of cells in the Barrett's training dataset. Small numbers of cells were seen on the Barrett's cell slides selected for cytopathology analysis: this was likely due to suboptimal cell preparation for the relatively smaller Barrett's cells (smaller than squamous or dysplastic cells, which tended to clump together on our cell preparation) – a difficulty of optimizing cell preparation and centrifugation for different cell types. Additionally, the small Barrett's cells were more likely to be missed by our cell detection algorithm (the binary mask). This left a low number of cells that were identified by the binary mask, and also identified by both reporting cytopathologists as unequivocally representing Barrett's cells.'

2. The authors should clearly state that this automated cytological detection is unsuitable for Barrett's esophagus screening because of low sensitivity.

The following sentence has been added to the Conclusions:

'The low sensitivity for non-dysplastic Barrett's using the present technique is not suitable as for screening purposes.'

3. What diagnosis was obtained by FITR spectroscopy in the case of misdiagnosis?

We have inserted the following table as part of Table 4. This confusion matrix shows the diagnosis in cases of misclassification.

	Predicted pathology			
True Pathology	**NSQ**	**BE**	**DYS/AC**	**TOTALS**

NSQ	**15**	0	4	19
BE	7	**10**	15	32
DYS/AC	7	0	**35**	42
TOTALS	29	10	44	93

4. Higher quality of figure 3 is required.

We have reproduced this figure at higher resolution (see figure file uploaded).

AUTOMATED CYTOLOGICAL DETECTION OF BARRETT'S NEOPLASIA WITH INFRARED SPECTROSCOPY

O. Old[1,2], G. Lloyd[1,3], M. Isabelle[1], L. M. Almond[1,4], C. Kendall[1,3], K. Baxter[5], N. Shepherd[5], A.C. Shore[6], N. Stone[3], H. Barr[1,5]

[1]Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucester, GL1 3NN, UK.

[2]University of Exeter Medical School, Royal Devon and Exeter NHS Foundation Trust, Exeter, EX2 5DW, UK.

[3]School of Physics and Astronomy, University of Exeter, Exeter, EX4 4QL, UK.

[4]Heartlands Hospital, Bordesley Green East, Birmingham, B9 5SS, UK

[5]Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucester, GL1 3NN, UK.

[6]NIHR Exeter Clinical Research Facility, University of Exeter Medical School, Royal Devon and Exeter NHS Foundation Trust, Exeter, EX2 5DW, UK.

Short title: Detection of BE with FTIR cytology

Word count: 4944

Abstract

Background

Development of a non-endoscopic test for Barrett's esophagus (BE) would revolutionize population screening and surveillance for patients with BE. Swallowed cell collection devices have recently been developed to obtain cytology from the esophagus: automated detection of neoplasia in such samples would enable large-scale screening and surveillance.

Methods

Fourier Transform infrared spectroscopy (FTIR) was used to develop an automated tool for detection of BE and Barrett's neoplasia in esophageal cell samples. Cytology brushings were collected at endoscopy, cytospun onto slides and FTIR images measured. An automated cell recognition program was developed to identify individual cells on the slide.

Results

Cytology review and contemporaneous histology was used to inform a training dataset containing 141 cells from 17 patients. A classification model was constructed using principal component analysis (PCA) fed linear discriminant analysis (LDA), then tested using leave one sample out cross validation (LOSOCV). Applying this training model to whole slide samples, a threshold voting system was used to classify samples according to their constituent cells. Across the entire dataset of 115 FTIR maps from 66 patients, whole samples were classified with sensitivity and specificity respectively as follows: normal squamous 79.0% and 81.1%, non-dysplastic Barrett's 31.3% and 100%, and neoplastic Barrett's 83.3% and 62.7%.

Conclusions

Analysis of esophageal cell samples can be performed using FTIR with reasonable

sensitivity for Barrett’s neoplasia, though poor specificity with the current technique.

Keywords:

Barrett's esophagus; Esophageal adenocarcinoma; Screening; Cytology; FTIR spectroscopy

BACKGROUND

The current British Society of Gastroenterology (BSG) guidelines list several future developments that would 'revolutionize the care of individuals with Barrett's esophagus and should be priorities for policy makers and funders', of which the number one item listed is 'a non-endoscopic test(s) for diagnosis and surveillance' [1]. Development of an accurate, minimally invasive, relatively low-cost test could radically alter current models of endoscopic surveillance, and overcome the greatest obstacles to screening for Barrett's, namely the cost, acceptability and risks of endoscopy.

Swallowed cell collection devices offer a potential non-endoscopic means of sampling from the esophagus. Balloon collection devices have been used in eastern Asia for several decades as a screening tool for squamous cell carcinoma of the esophagus [2–4]. A balloon device has been trialed for Barrett's-associated neoplasia, but challenges included inadequate cell collection and reduced sensitivity for dysplasia [5]. Results from the same study suggested that brush cytology has the potential to collect a representative sample, but the difficulty of interpretation of low grade dysplasia may limit the sensitivity using conventional cytological assessment.

If a swallowed cell collection device were to be used as a screening tool for Barrett's this would pose a number of challenges for conventional cytological assessment. Firstly, esophageal cytology is performed relatively infrequently, and expertise in this field is correspondingly limited. Secondly, assessment of cells may be challenging and shows variable correlation with histology taken contemporaneously [6–8]. Thirdly, a swallowed device that enters the stomach may have glandular cells from the stomach and these must be differentiated from esophageal glandular metaplasia.

Fourthly, if expert cytological assessment were required for every sample collected as part of population screening, this would require significant resources.

The recently developed Cytosponge™ device aims to overcome a number of these challenges through immunostaining with Trefoil Factor 3 (TFF3), with promising results for detection of BE [9,10].

An alternative approach to conventional cytological cell classification uses Fourier Transform infrared spectroscopy (FTIR). In FTIR spectroscopy, infrared radiation is passed through a sample. Some of the infrared radiation is absorbed by the sample and some of it passes through (i.e. it is transmitted). The resulting spectrum represents the molecular absorption and transmission, creating a molecular fingerprint of the sample based on its biomolecular composition. This makes infrared spectroscopy useful for several types of analysis.

Subtle variations in sample biochemistry can be detected, highlighting different pathology states, and FTIR spectra can be used to build classification models to assign pathology labels to a range of biomedical samples [11].

To date there have been a small number of studies showing the ability of FTIR spectroscopy to classify BE using esophageal tissue [12–14], and only one small study showing feasibility in detection of BE using esophageal cells [15]. The present study was designed as a larger study to evaluate FTIR as a means for detecting Barrett's neoplasia in esophageal cells from brush cytology samples.

METHODS

Sample collection

Although the long term goal is non-endoscopic cell collection, samples in this study were taken under direct vision at endoscopy to allow contemporaneous biopsy and rigorous inclusion criteria to inform a training dataset.

Ethical approval was obtained for this study and all patients participating in the study provided informed consent. Samples were collected from patients undergoing scheduled endoscopy for Barrett's surveillance. In order to enrich the sample population for patients with esophageal adenocarcinoma, some samples were also collected from patients undergoing surgery for esophageal cancer. Normal esophageal squamous cells, used as controls, were collected from patients undergoing routine endoscopy in whom no endoscopic or histological abnormality was identified.

Cytology samples were collected at endoscopy using an endoscopic cytology brush passed down the instrument channel of an endoscope, under direct vision at endoscopy. The cytology brush containing the cells was then stored in formalin at room temperature until slide preparation.

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

For later analysis of whole samples included in the test dataset, the classification of the whole sample was based on the endoscopy and biopsy result as the gold

standard. To remove further the possibility of misclassification, patients were only included in the test dataset if they had no history of more advanced disease i.e. to be included as a normal control they must have no history of BE, and to be included as a BE case they must have no history of esophageal dysplasia/adenocarcinoma.

Sample measurement

Cell samples were transferred to calcium fluoride slides for measurement; a detailed protocol of this procedure is included in Appendix A of the Electronic Supplementary Material.

Samples were measured at the Biophotonics Unit at Gloucestershire Royal Hospital using a Perkin Elmer Spectrum One FTIR spectrometer with a Perkin Elmer Spotlight 400 imaging system. Infrared absorption maps were obtained in transmission mode, raster scanning at 6.25μm per pixel, using 4cm^{-1} spectral resolution across a wavenumber range of 750-4000cm^{-1}. An initial background reading was taken using 120 scans per pixel from an acellular region of the slide. Maps of infrared absorption were collected across a 4mm x 2mm region of each slide, measuring 2 scans per pixel.

Data processing: automated identification of cells

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

Initial data processing steps were applied to enhance signal to noise ratios in the measured cell spectra, and prepare spectra for further analysis.

A post-processing algorithm was developed to detect cellular regions and extract the spectra from this region, assigning it to a particular cell. The spectra measured from each individual cell could then be analyzed and a classification label assigned to that particular cell. The pre-processing steps and post-processing algorithm are described further in Appendix B of the Electronic Supplementary Material. Similar techniques have been described previously by a number of authors [16–18].

Developing a training dataset

The training dataset was composed of cells from a preliminary group of patients. After identification by the binary mask algorithm, cells whose classification on appearance by two cytopathologists was in agreement with contemporaneous endoscopy and biopsy from the same region were included. The dysplasia/adenocarcinoma group included any degree of dysplasia (low or high grade), but samples classified as ‘indefinite for dysplasia’ were excluded.

Multivariate analysis was then performed to identify spectral differences between cells from the different pathology groups to enable classification based on cell spectra. Principal component analysis (PCA) was used to identify variance between the spectra and reduce the complexity of the dataset. Linear discriminant analysis (LDA) was then used to incorporate information about which pathology groups the spectra came from, and build diagnostic classification models. Diagnostic classification models were then tested using leave-one-sample-out-cross validation

(LOSOCV): each sample is sequentially tested in training models that contains data from every sample except the one being tested.

Classification models developed using the training dataset could then be applied to the full dataset, to classify whole samples from individual patients. Cells that were included in the training dataset were excluded from the test set, to prevent overestimation of performance.

Since not every cell on a slide would necessarily have the same classification, a 'threshold voting' system was used to assign a classification to a given sample. For example, a 30% threshold per cell would mean that, if 30% or more spectra from that cell were classified as dysplastic, the entire cell is classified as dysplastic. Similarly, if ≥30% of cells on a slide are found to be dysplastic the sample is classified as dysplastic. This is a higher threshold than that used by cytologists, whereby the presence of any dysplasia is sufficient for classification as dysplastic, but aimed to ensure higher specificity when very large numbers of cells were being tested for each patient.

RESULTS

Training dataset

The training dataset was developed using 141 cells from 17 patients (8 normal squamous, 4 BE, 5 dysplasia/adenocarcinoma), with a total of 1,480 spectra. The data included in the training dataset are shown in Table 1.

A classification model was constructed using PCA-fed LDA. The linear discriminant functions are shown in Figure 1. LDA achieves good grouping and separation of each of the pathology groups, but there remains a cluster of BE datapoints close to

the other two groups. Every one of these outlying datapoints was from a single sample.

Table 1 Summary of data included in the training dataset.

	No. of patients	No. of cell regions	No. of spectra for analysis
NSQ	8	22	726
BE	4	12	76
DYS/AC	5	22	678
TOTAL	17	56	1,480

NSQ normal squamous; BE Barrett's esophagus; DYS/AC dysplasia/adenocarcinoma

TAKE IN FIGURE 1

Figure 1 Scatter plot of all spectra in training dataset, plotted by linear discriminant (LD) function. NSQ = normal squamous, Barrett's = Barrett's esophagus, Adenoca. = adenocarcinoma.

The performance of the training dataset classification model was then tested using LOSOCV: the results are shown in Table 2.

Table 2 Performance of the training cells 3 group classification model

	NSQ	BE	DYS/AC
Sensitivity % (SD)	83.6 (5.2)	62.8 (0.8)	69.5 (4.5)
Specificity % (SD)	70.8 (3.8)	97.5 (0.9)	87.8 (3.5)

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

The sensitivity of the model for individual spectra is reasonably good for normal squamous cells at 83.6%, but only moderate for the Barrett's and dysplasia/adenocarcinoma cells. The standard deviation was largest for the normal squamous cells, as would be expected since the cells chosen were different for each iteration. The size of the variation in the sensitivity for the dysplasia/adenocarcinoma group reflects the overlap with the normal squamous cells and hence this varies depending on which cells are included. There was very little variation in the Barrett's result, with a small standard deviation.

Spectral differences between pathology groups

The mean spectra from each of the three pathology groups are shown in Figure 2 below. The spectra are presented as second derivatives in order to clarify the position of spectral peaks and highlight differences between the groups.

TAKE IN FIGURE 2

Figure 2 Second derivative of the mean spectra from A) NSQ B) BE C) DYS/AC, with the text color indicating the likely biomolecule mainly responsible for that peak. NSQ normal squamous; BE Barrett's esophagus; DYS/AC dysplasia/adenocarcinoma. Green = glycogen, blue = glycoprotein, red = DNA, Black = amino acid/protein, Purple = mixed contributions/multiple possibilities.

Assigning biochemical labels to spectral peaks is tentative, as there are frequently multiple possible bond vibrations that can give a spectral peak at a given wavenumber. However, based on previous studies in esophageal tissue [12–14] and previous work by the group in esophageal tissue [19], there are a number of important differences seen between the pathology groups in this study.

The BE and dysplasia/adenocarcinoma groups have an additional spectral peak at 969cm^{-1} that probably represents DNA, and may reflect a higher DNA content in these cell types.

At one of the key glycoprotein regions around 1080cm^{-1}, the BE group shows a doublet with peaks at 1067 and 1084cm^{-1}, whereas the normal squamous and dysplasia/adenocarcinoma groups have a single peak at 1077cm^{-1}. This may be due to differences/increase in mucin content within the BE cells.

There are three peaks corresponding to glycogen in the normal squamous cells at 995, 1024 and 1153cm^{-1}: these peaks are either smaller or show different configurations in the BE and dysplasia/adenocarcinoma groups, which may reflect a lower glycogen content in these cell, as might be expected with their pathological state.

Test dataset

The classification model developed from the training dataset was then applied to the full dataset. The steps involved are illustrated in Figure 3: first the binary mask algorithm was used to identify individual cells on each slide, then the classification model applied to assign a pathology label to each cell on the slide based on its FTIR spectra, then an overall pathology classification given to each slide. This predicted pathology was compared against the biopsy result from the same region and sensitivity and specificity calculated using this gold standard.

TAKE IN FIGURE 3

Figure 3 Steps involved in assigning pathology labels to every cell A) unstained slide, B) binary mask identifies cell regions, C) training model applied to cells to assign pathology label.

The number of samples and spectra included in the test dataset (after the same pre-processing steps as those applied to the training dataset) are shown in Table 3.

Table 3 Total number of samples included in the test dataset

	No. of patients	No. of FTIR maps	No. of cell regions	No. of spectra for analysis
NSQ	18	19	722	27,662
BE	21	32	1,891	48,322
DYS/AC	25	42	2,620	54,418
TOTAL	64	93	5,233	130,402

NSQ normal squamous; BE Barrett's esophagus; DYS/AC dysplasia/ adenocarcinoma.

The results from application of the training model to this test dataset are shown in Table 4.

Table 4 Classification performance of the training model applied to the whole samples as a test dataset. A voting threshold was used for classifying each individual cell on a slide (30% threshold) and for classifying the sample overall (30% threshold). Table 4A shows the sensitivity and specificity, whilst Table 4B shows the confusion matrix, showing the prediction for every sample. NSQ normal squamous; BE Barrett's esophagus; DYS/AC dysplasia/ adenocarcinoma.

A

	NSQ	**BE**	**DYS/AC**
Sensitivity %	79.0	31.3	83.3
Specificity %	81.1	100	62.7

B

	Predicted pathology			
True Pathology	**NSQ**	**BE**	**DYS/AC**	**TOTALS**
NSQ	**15**	0	4	19
BE	7	**10**	15	32
DYS/AC	7	0	**35**	42
TOTALS	29	10	44	93

Overall, the detection of normal squamous and dysplasia/adenocarcinoma samples was reasonably good, with a sensitivity of 79.0% and 83.3% respectively. Specificity for dysplasia/adenocarcinoma was low at 62.7%, and detection of Barrett's was poor, with only 31.3% of samples classified correctly.

DISCUSSION

FTIR was investigated as a means of identifying Barrett's esophagus and associated neoplasia in esophageal cells, for potential use in conjunction with a non-endoscopic cell collection device. In this study, a good classification performance was seen for normal squamous samples (sensitivity 79.0%), and dysplasia/ adenocarcinoma (83.3%), but the identification of BE samples was poor (sensitivity 31.3%).

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

Certain key differences were seen between the mean spectra from the different pathology groups, that were consistent with possible biochemical differences between the cells. These included a higher glycogen content in normal squamous cells, altered mucin content in BE cells, and higher DNA content in both BE and dysplastic/ adenocarcinoma cells compared to normal squamous cells.

Nonetheless there are several reasons why the results for the BE cells may have been poor. The major limiting factor in this study was the small number of cells in the Barrett's training dataset. Small numbers of cells were seen on the Barrett's cell slides selected for cytopathology analysis: this was likely due to suboptimal cell

preparation for the relatively smaller Barrett's cells (smaller than squamous or dysplastic cells, which tended to clump together on our cell preparation) – a difficulty of optimizing cell preparation and centrifugation for different cell types. Additionally, the small Barrett's cells were more likely to be missed by our cell detection algorithm (the binary mask). This left a low number of cells that were identified by the binary mask, and also identified by both reporting cytopathologists as unequivocally representing Barrett's cells.

Not only did the Barrett's group have the smallest number of cells, but these tended to be small, isolated cells (rather than a cluster of cells of the same type), and thus contained very few total spectra (76 in this group, versus 678 in the dysplasia group, and 726 in the normal squamous). The training model is therefore skewed away from classifying spectra into the Barrett's group, and the Barrett's group can also more readily be affected by a small number of outlying spectra. This makes the Barrett's dataset prone to strong influence from differences due to individual samples or patients rather than true biochemical differences due to pathology.

High numbers of squamous cells were noted on many samples taken only from glandular regions of esophagus. It is possible that this resulted from sampling only the most superficial cells, which may include squamous cells that originated more proximally in the esophagus that have sloughed off and been deposited more distally over a glandular region.

Given the presence of squamous cells seen on many slides, the relative homogeneity of the results is perhaps surprising. Whilst it is possible that artefact in the form of 'between-patient' differences (as opposed to true pathological differences) contribute to this, the small number of patients in the training model relative to the test dataset makes this explanation insufficient. A further possibility is

that biochemical changes precedes a morphological change in cell appearance. If a region of dysplasia or adenocarcinoma underwent a field change (probably reflecting genetic change) that preceded phenotypic change in the cells, there may be a detectable biochemical difference in cells that appear squamous. Thus spectroscopy may be able to provide insights into biochemical changes not detectable with conventional microscopy.

Interestingly, a similar finding was reported in the largest study to date of IR spectroscopy for cell classification in cervical samples: Gajjar et al. [20] found that their results correlated poorly with conventional cytology, but showed better correlation with contemporaneous histology from the same region. This paper cites poor cytology sensitivity and specificity as the reason for this, but an analogous explanation of field change that has not occurred in every cell (and again, particularly in the superficial cells), is another possibility.

This finding is corroborated by the only previous study of FTIR using esophageal cells [15], in which cells that appeared squamous classified according to the underlying tissue histology. This consistent finding therefore supports the theory that genetic and biochemical change may precede morphological change in some cells.

In this small study of 10 samples, Townsend et al. reported very high accuracy for classification of Barrett's and dysplasia. They reported a sensitivity to detect Barrett's versus normal squamous of 95.5%, normal squamous versus dysplasia 93.4%, and Barrett's versus dysplasia 88.7%. However, spectra from the same patients were included in both test and training datasets, and since these results are drawn from a very small number of patients it is possible that 'between patient' differences (as opposed to 'between pathology' differences) contributed to this result.

The discriminatory spectral features seen in the work by Townsend et al. are similar to the findings seen in the inverse second derivative spectra in our study and also correlate with the findings from our tissue mapping study [19] and previous FTIR work in the esophagus [12–14]. For example, the amide I peak at around 1650cm^{-1} is strongest in the normal squamous group, the DNA peak at 1235cm^{-1} is strongest in the dysplastic cells, and the glycogen peak at 1020cm^{-1} is strongest in the normal squamous cells and almost absent in the dysplasia group. This strengthens the findings in their study, and lends further weight to an argument for an underlying biochemical difference between the pathology groups, as seen in our analysis of the inverse second derivative spectra.

There have been a small number of other studies using FTIR to classify cell pathology with similar methodology. These have shown the ability to classify squamous cell samples from the cervix, urinary tract and head and neck [16,17,20,21], although this work focused on proof of concept and spectral differences, and did not publish equivalent figures for sensitivity and specificity.

The first trial data (BEST2) from use of the Cytosponge™ found overall sensitivity for detecting Barrett's 79.9%, with this figure increasing to 87.2% in those with ≥3cm of circumferential Barrett's, and a specificity of 92.4% [9]. This is comparable to the sensitivity seen in our study for normal squamous (79.0%) or dysplasia/adenocarcinoma (83.3%). The BEST2 study did not attempt to discriminate dysplasia, but there are plans for future work to incorporate risk stratification using DNA analysis for p53 mutations.

Identification of Barrett's/dysplasia in combination is sufficient if used solely as a screening tool to identify those who require endoscopy. However, reliable detection

of dysplasia could potentially replace endoscopic surveillance. The results of our study suggest that this may be achievable with FTIR spectral analysis of cells. However, this is a multistep technique which requires further optimization and validation in combination with a non-endoscopic collection device such as Cytosponge™.

The major source of variability in this study was the widely varying cell density in slide samples. Further optimization of techniques for brushing, fixative and centrifugation may improve this [22].

There is much potential to refine the binary mask with a combination of further spectral information and size criteria to give a highly accurate cell identification tool. One approach could use a spectral marker of DNA (e.g. the 1234cm^{-1} peak) to identify cell nuclei, in combination with associated cell cytoplasm (represented by the 1650cm^{-1} peak). This could be used in combination with size criteria to quantify the amount of nuclear material present.

A two-stage analysis may be developed whereby cells are initially separated into squamous or columnar by a predictive model, and then a further analysis used to separate dysplastic and non-dysplastic cells. A two-stage model was attempted using the current dataset, giving very similar results to the 3-group model presented here (two-stage results not shown).

Whilst this study used samples collected at endoscopy in order to enable histological and endoscopic validation, there may be further challenges using a non-endoscopic cell collection device that collects from the stomach and entire esophagus. The issue of differentiating pathological glandular cells from stomach cells was not examined in the present study.

Another obstacle to clinical implementation is the time needed for sample measurement. If a faster automated cell detection process could be used prior to infrared measurement, this would avoid measuring large regions that do not contain cells, and focus solely on collecting useful spectra.

Cytopathology review is a further potential source of error. In this study it was performed by two cytopathologists together, not independently. Additionally, cytopathologists are not used to reviewing single cells in isolation, but more usually look at a whole sample for assessment. Presentation of cells in isolation could mean the decision is affected by variable staining between slides, which might be accounted for by taking the slide as a whole.

One of the major obstacles to future work in this area is obtaining a reliable gold standard against which to test the training model. The poor sensitivity and specificity in comparable fields (e.g. cervical cytology) suggests that this may limit the usefulness of cytology as a gold standard. Comparing against histology (as in the whole sample test dataset) may be more accurate.

Although the relatively poor sensitivity and specificity of cytology causes problems for testing, this supports an argument for developing spectral cytopathology since this an area in which diagnostic performance could readily be improved. This technology could potentially be applied in a range of different pathologies and organ systems.

Conclusions

FTIR offers a potential automated method of identifying Barrett's neoplasia in esophageal cell samples. Accurate identification of neoplasia could augment or

replace current models of endoscopic surveillance. High sensitivity for neoplasia would be a key feature if used in clinical practice, but further work is needed to optimize the current technique and improve specificity prior to clinical translation. The low sensitivity for non-dysplastic Barrett's using the present technique is not suitable for screening purposes.

Acknowledgements:

The authors would like to thank Doug Townsend and Max Diem from Northeastern University, Boston, for all their advice on many technical aspects of spectral cytopathology.

Oliver Old was in receipt of a Royal College of Surgeons of England Surgical Research Fellowship during this study.

References

1. Fitzgerald RC, di Pietro M, Ragunath K, Ang Y, Kang J-Y, Watson P, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. Gut. 2014;63:7–42.

2. Shu YJ. Cytopathology of the esophagus. An overview of esophageal cytopathology in China. Acta Cytol. 1983;27:7–16.

3. Dawsey SM, Shen Q, Nieberg RK, Liu SF, English SA, Cao J, et al. Studies of esophageal balloon cytology in Linxian, China. Cancer Epidemiol. Biomarkers Prev. 1997;6:121–30.

4. Spechler S. Barrett's esophagus: Should we brush off this ballooning problem? Gastroenterology. 1997;112:2138–42.

5. Falk GW, Chittajallu R, Goldblum JR, Biscotti C V, Geisinger KIMR, Petras RE, et al. Surveillance of Patients With Barrett ' s Esophagus for. 1997;1787–97.

6. Hughes JH, Ph D, Cohen MB. Is the Cytologic Diagnosis of Esophageal Glandular Dysplasia Feasible ? 1998;18:312–6.

7. Geisinger KR, Teot LA, Richter JE. A comparative cytopathologic and histologic study of atypia, dysplasia, and adenocarcinoma in Barrett's esophagus. Cancer. 1992;69:8–16.

8. Hardwick RH, Morgan RJ, Warren BF, Lott M, Alderson D. Brush cytology in the diagnosis of neoplasia in Barrett's esophagus. Dis. Esophagus. 1997;10:233–7.

9. Ross-Innes CS, Debiram-Beecham I, O'Donovan M, Walker E, Varghese S, Lao-Sirieix P, et al. Evaluation of a Minimally Invasive Cell Sampling Device Coupled with Assessment of Trefoil Factor 3 Expression for Diagnosing Barrett's Esophagus: A Multi-Center Case–Control Study. Franco EL, editor. PLOS Med. 2015;12:e1001780.

10. Kadri SR, Lao-Sirieix P, O'Donovan M, Debiram I, Das M, Blazeby JM, et al. Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. BMJ. 2010;341:c4372.

11. Old O, Fullwood L, Scott R, Lloyd G, Almond L, NShepherd N, et al. Vibrational Spectroscopy for cancer diagnostics. Anal. methods. 2014;6:3901–17.

12. Wang TD, Triadafilopoulos G, Crawford JM, Dixon LR, Bhandari T, Sahbaie P, et al. Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2007;104:15864–9.

13. Quaroni L, Casson AG. Characterization of Barrett esophagus and esophageal adenocarcinoma by Fourier-transform infrared microscopy. Analyst. 2009;134:1240–6.

14. Amrania H, Antonacci G, Chan C-H, Drummond L, Otto WR, Wright N a, et al. Digistain: a digital staining instrument for histopathology. Opt. Express. 2012;20:7290–9.

15. Townsend D, Miljković M, Bird B, Lenau K, Old O, Almond M, et al. Infrared micro-spectroscopy for cyto-pathological classification of esophageal cells. Analyst. 2015;140:2215–23.

16. Schubert JM, Bird B, Papamarkakis K, Miljković M, Bedrossian K, Laver N, et al. Spectral cytopathology of cervical samples: detecting cellular abnormalities in cytologically normal cells. Lab. Invest. 2010;90:1068–77.

17. Papamarkakis K, Bird B, Schubert JM, Miljković M, Wein R, Bedrossian K, et al. Cytopathology by optical methods: spectral cytopathology of the oral mucosa. Lab. Invest. 2010;90:589–98.

18. Miljković M, Bird B, Lenau K, Mazur AI, Diem M. Spectral cytopathology: new aspects of data collection, manipulation and confounding effects. Analyst.

2013;138:3975–82.

19. Old OJ, Lloyd GR, Nallala J, Isabelle M, Almond LM, Shepherd NA, et al. Rapid infrared mapping for highly accurate automated histology in Barrett's oesophagus. Analyst. 2016;

20. Gajjar K, Ahmadzai A a, Valasoulis G, Trevisan J, Founta C, Nasioutziki M, et al. Histology verification demonstrates that biospectroscopy analysis of cervical cytology identifies underlying disease more accurately than conventional screening: removing the confounder of discordance. PLoS One. 2014;9:e82416.

21. Bird B, Romeo MJ, Diem M, Bedrossian K, Laver N, Naber S. Cytology by Infrared Micro-Spectroscopy: Automatic Distinction of Cell Types in Urinary Cytology. Vib. Spectrosc. 2008;48:101–6.

22. Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, et al. Using Fourier transform IR spectroscopy to analyze biological materials. Nat. Protoc. 2014;9:1771–91.

23. Bassan P, Byrne HJ, Bonnier F, Lee J, Dumas P, Gardner P. Resonant Mie scattering in infrared spectroscopy of biological materials--understanding the "dispersion artefact". Analyst. 2009;134:1586–93.

24. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, et al. Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples. Analyst. 2010;135:268–77.

Figure 1

Click here to download Figure Figure 1 v2.pdf

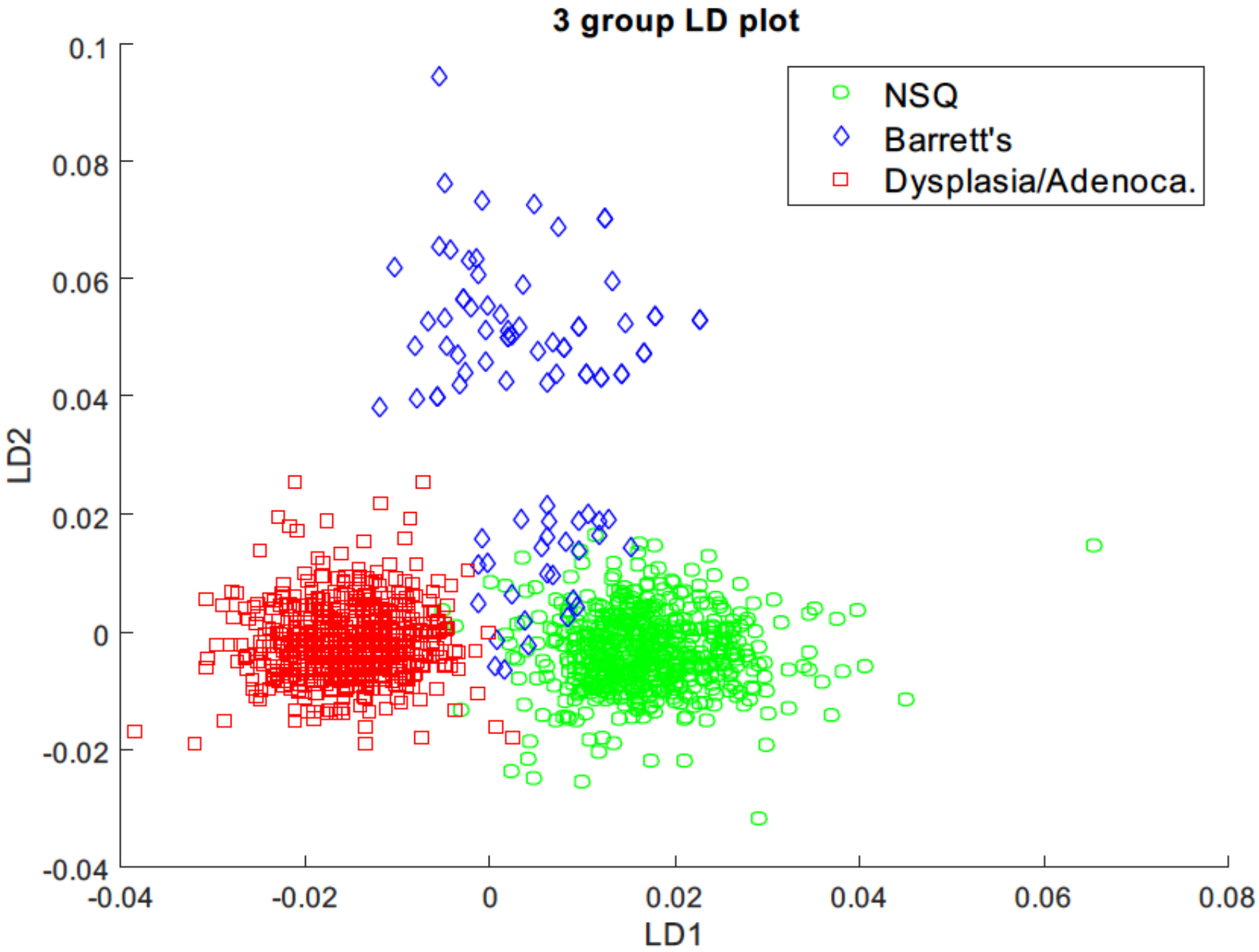
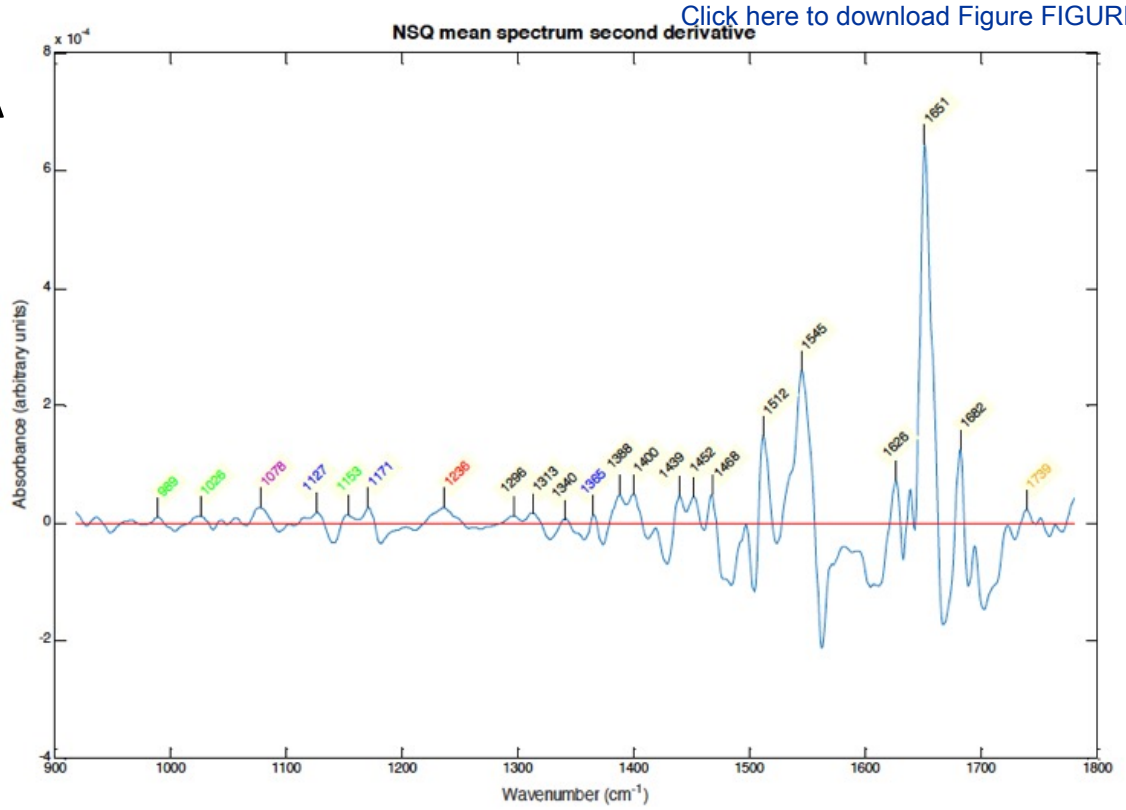


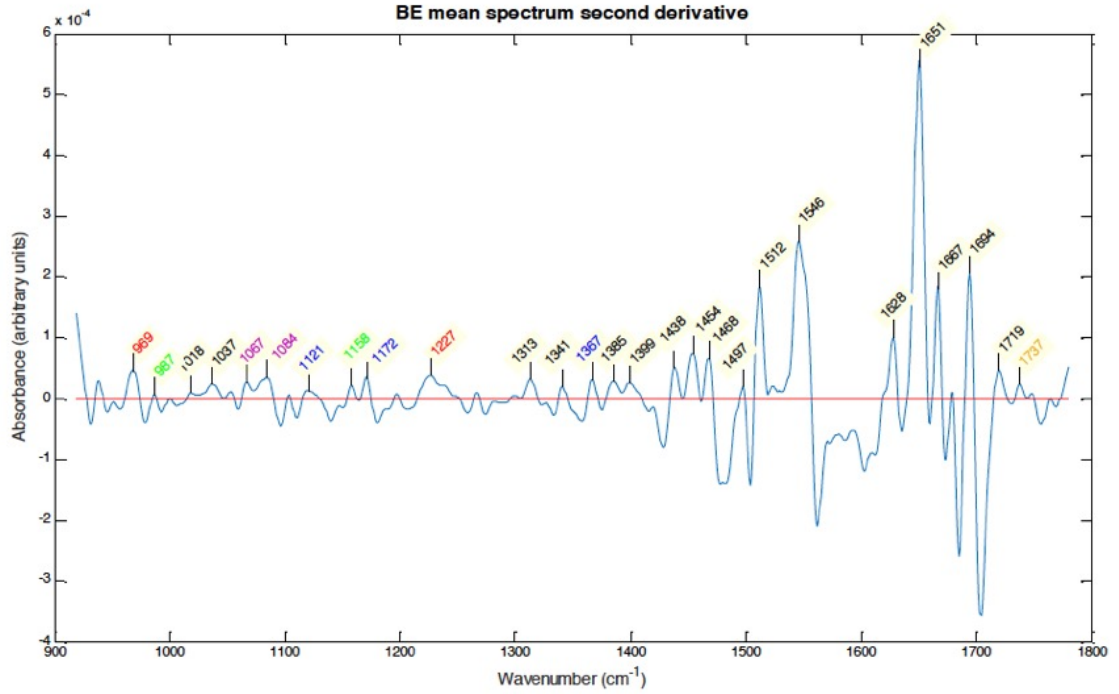
Figure 2

Click here to download Figure FIGURE 2.pdf

A



B



C

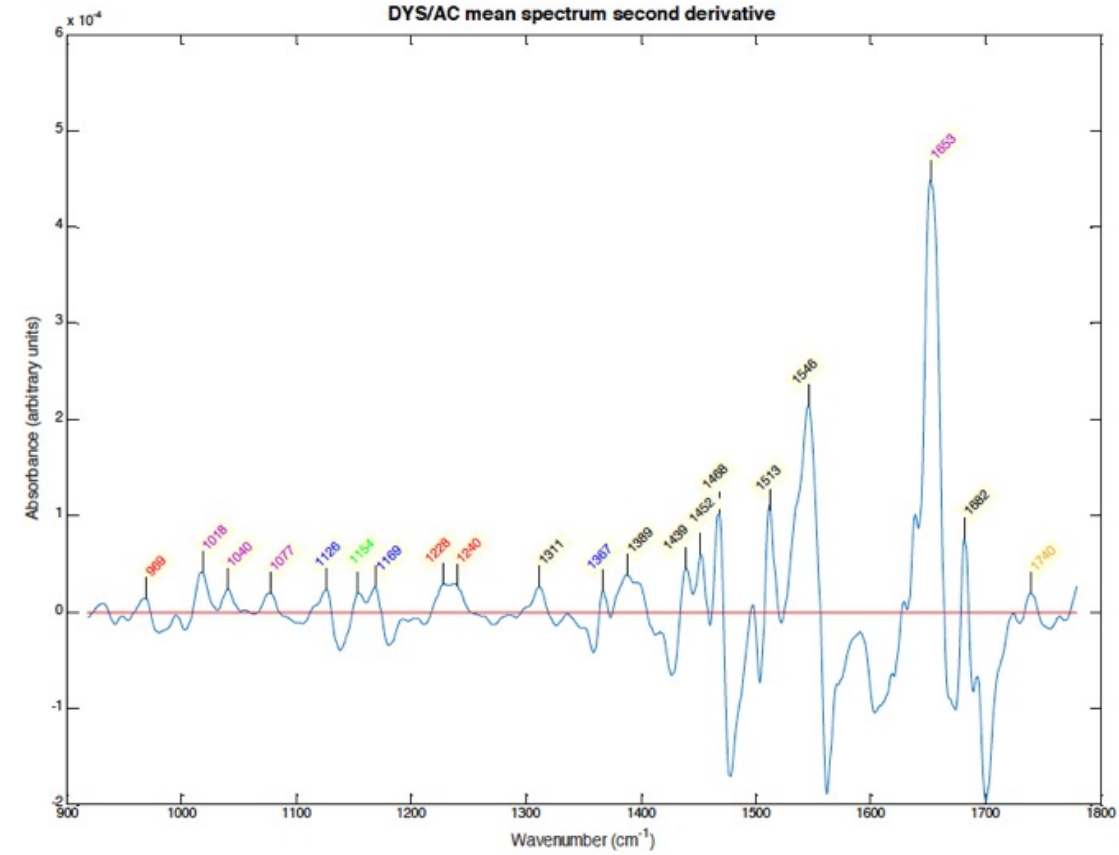


Figure 3

Click here to download Figure Revised Figure 3.pdf

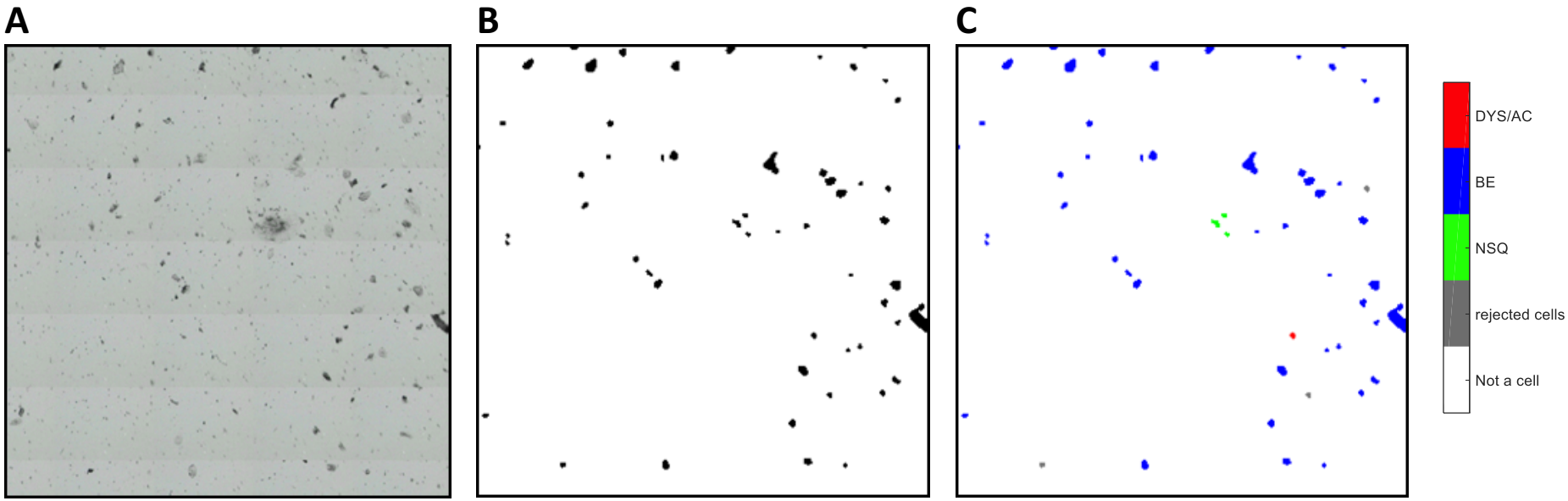


Figure 4 Supplementary Material (Electric Supplementary Material)

Click here to download Conflict of Interest Disclosure Statement COI form.docx

The Japanese Society of Gastroenterology (JSGE)

Journal of Gastroenterology

Conflict of Interest Disclosure Statement Form

The corresponding author should upload the form online.

When submitting a manuscript to the Journal of Gastroenterology, all authors are required to disclose any financial relationship (**within the last 2 years**) with a biotechnology manufacturer, a pharmaceutical company, or other commercial entity that has an interest in the subject matter or materials discussed in the manuscript. The matters requiring disclosure are outlined in the JSGE Conflict of Interest Policy.

Disclosed Potential Conflict of Interest

1. Employment/Leadership position/ Advisory role (1,000,000 yen or more)
2. Stock ownership or options (Profit of 1,000,000 yen or more/ownership of 5% or more of total shares)
3. Patent royalties/licensing fees (1,000,000 yen or more)
4. Honoraria (e.g. lecture fees) (1,000,000 yen or more)
5. Fees for promotional materials (e.g. manuscript fee) (1,000,000 yen or more)
6. Commercial research funding (2,000,000 yen or more)
7. Others (e.g. trips, travel, or gifts, which are not related to research, education and medical practice) (50,000 yen or more)

If any of the above items (1 to 7) apply to author(s) of the article, the corresponding author should provide the statement in the space below by using the following examples for each author.
"A (author name) received a research grant from Z; B serves as a consultant to Y (entity name); C received lecture fees from X; D received honoraria for writing promotional material for W;E holds a patent on V; F's spouse is chairman of U."

If none of the authors have a relationship matching the above listed items (1 to 7), please provide the statement: "The authors declare that they have no conflict of interest" in the space below.

The authors declare that they have no conflict of interest.

When your manuscript is accepted for publication, all of the disclosures will appear in your article as a "Conflict of Interest Statement" in the Journal of Gastroenterology.