Central Lancashire Online Knowledge (CLoK)

Attenuated total reflection Fourier-transform infrared spectral discrimination in human bodily fluids of oesophageal transformation to adenocarcinoma

Ishaan Maitra,[a*] Camilo L. M. Morais,[a] Kássio M. G. Lima,[a,b*] Katherine M. Ashton,[c] Ravindra S. Date[c] and Francis L. Martin[a*]

[a]*School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, UK*

[b]*Institute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal, Brazil*

[c]*Lancashire Teaching Hospitals NHS Foundation Trust, Royal Preston Hospital, Preston PR2 9HT, UK*

*ishaan.maitra@googlemail.com; kassiolima@gmail.com; flmartin@uclan.ac.uk

Abstract

Diagnostic tools for the detection of early-stage oesophageal adenocarcinoma (OAC) are urgently needed. Our aim was to develop an accurate and inexpensive method using biofluids (plasma, serum, saliva or urine) for detecting oesophageal stages through to OAC (squamous; inflammatory; Barrett's; low-grade dysplasia; high-grade dysplasia; OAC) using attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy. ATR-FTIR spectroscopy coupled with variable selection methods, with successive projections or genetic algorithms (GA) combined with quadratic discriminant analysis (QDA) were employed to identify spectral biomarkers in biofluids for accurate diagnosis in a hospital setting of different stages through to OAC. Quality metrics (Accuracy, Sensitivity, Specificity and F-score) and biomarkers of disease were computed for each model. For plasma, GA-QDA models using 15 wavenumbers achieved 100% classification for four classes. For saliva, PCA-QDA models achieved 100% for the inflammatory stage and high-quality metrics for other classes. For serum, GA-QDA models achieved 100% performance for the OAC stage using 13 wavenumbers. For urine, PCA-QDA models achieved 100% performance for all classes. Selected wavenumbers using a Student's t-test (95% confidence interval) identified a differentiation of the stages on each biofluid: plasma (929 cm^{-1} to 1,431 cm^{-1}, associated with DNA/RNA and proteins); saliva (1,000 cm^{-1} to 1,150 cm^{-1}, associated with DNA/RNA region); serum (1,435 cm^{-1} to 1,573 cm^{-1}, associated with methyl groups of proteins and Amide II absorption); and, urine (1,681 cm^{-1} to 1777 cm^{-1}, associated with a high frequency vibration of an antiparallel β-sheet of Amide I and stretching vibration of lipids). Our methods have demonstrated excellent efficacy for a rapid, cost-effective method of diagnosis for specific stages to OAC. These findings suggest a potential diagnostic tool for oesophageal cancer and could be translated into clinical practice.

Keywords: ATR-FTIR spectroscopy; Biofluids; Classification techniques; Diagnosis; Oesophageal cancer; Variable selection methods

Introduction

The rising incidence of oesophageal cancer over the past three decades coincides with a change in histologic type and primary tumour location.[1] The incidence of oesophageal adenocarcinoma (OAC) in the Western world, which is a long-term complication of damage by gastro-oesophageal reflux, has been rising over recent decades.[2] The mounting incidence of obesity, which encourages gastro-oesophageal reflux appears to be a key factor. The second factor is the reduced incidence of *H. pylori* infection and associated atrophic gastritis. In turn, these factors decrease the acidity and peptic activity of gastric juice, damaging oesophageal mucosa.[2]

There is a proven association between Barrett's oesophagus (BO) and OAC as a result of chronic inflammation from gastro-oesophageal reflux disease (GORD).[3] Metaplastic cells in BO progress to OAC through intermediate histological stages of inflammation, low-grade dysplasia (LGD), high-grade dysplasia (HGD) and OAC. This transformation that takes several years.[4] OAC is aggressive and usually presents late with a poor prognosis.

BO is the only known precursor to OAC to date. GORD increases the risk of OAC by >40-fold compared with the general population.[4] There remains a lack of insight into its natural history.[5] Additionally, there are no reliable predictive biomarkers that might enable us to risk-stratify BO patients and identify those who would benefit most from endoscopic management.[6] Finally, prospective studies have not established a clear survival benefit for screening and surveillance in BO.[7] There are several guidelines regarding screening, surveillance and management of BO. All however fail to demonstrate clear evidence for an established benefit, cost-effectiveness and robust risk stratification for patients.[8]

A recent systematic review and meta-analysis conducted in 2017 confirmed that the risk of progression to HGD or OAC in BO patients is primarily determined by the presence or absence of LGD (OR 4.2 (2.1–8.5)).[9] The progression of BO to HGD or adenocarcinoma is further significantly increased with increasing length of BO segment [OR 1.2 (1.1–1.3) per additional cm in length]. Furthermore, Krishnamoorthi *et al.*[9] confirmed that older age and being male are risk factors for disease progression to HGD and OAC. Finally, the meta-analysis suggests that use of a proton pump inhibitor (OR 0.55 (95% CI 0.32; 0.96)) or statins (OR 0.48 (95% CI 0.31; 0.73)) significantly decreases the risk of progression to HGD or cancer.[9]

Early detection and prevention are key strategies to manage OAC. Early detection of cancer or dysplasia in BO allows intervention at an early stage. The argument as to which BO patients are most likely to benefit from surveillance and management hinges on the high prevalence of BO and the low cancer incidence among unselected BO cases *versus* the burden of invasive treatment and the high morbidity and mortality of OAC.[10] Methods investigated in the diagnosis of BO and establishment of LGD, HGD and OAC include high-resolution endoscopy,[11] chromoendoscopy,[12] autofluorescence,[13] narrow band imaging (NBI),[14] and optical coherence tomography (OCT)[15] amongst others.[16-18]

Although these methods have been widely used in the diagnosis and differential diagnosis of oesophageal diseases, two main drawbacks are noted. Current diagnostic methods rely on the collection of biopsy samples at endoscopy, which is an invasive and poor sampling technique despite rigorous protocols. The second drawback involves significant inter- and intra-observer variability for the endoscopist and pathologist. The

need for simple, reproducible and real-time information on the disease state by non-invasive methods has never been more relevant.

Attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy has been applied to detect various alterations in human tissues, cells or biofluids (urine, saliva, plasma, serum, cerebrospinal).[19-21] The derived ATR-FTIR spectrum is a series of wavenumber absorbance intensities, which correspond to particular frequencies, including Amide I (1650 cm^{-1}), Amide II (1550 cm^{-1}), Amide III (1260 cm^{-1}), carbohydrates (1155 cm^{-1}), glycogen (1030 cm^{-1}), lipids (1750 cm^{-1}), asymmetric phosphate stretching vibrations ($\nu_{as}PO_2^-$; 1225cm^{-1}), symmetric phosphate stretching vibrations ($\nu_s PO_2^-$; 1080 cm^{-1}) and protein phosphorylation (970 cm^{-1}).

ATR-FTIR spectroscopy has been used to identify neoplasia in ovary,[22] cervix,[23] breast,[24] brain,[25] prostate,[26] lung,[27] skin,[28] thyroid,[29] stomach,[30] colon[31] and pancreas.[32] Several groups have used IR spectroscopy for the detection of Barrett's oesophagus and pre-cancerous changes in oesophageal tissues.[33–36] To the best of our knowledge, ATR-FTIR spectroscopy has never been applied to interrogate the oesophageal stages of transformation to OAC, employing the biofluids plasma, serum, saliva or urine.

Herein we set out to develop an accurate, quick and inexpensive method using biofluids (plasma, saliva, serum or urine) towards detecting oesophageal stages through to OAC (normal; inflammatory; Barrett's; low-grade dysplasia (LGD); high-grade dysplasia (HGD); and, OAC). This was using a derived FTIR spectral region, or combination of variables, that reflects a specific biochemical feature of disease in human bodily fluids. We employed successive projections algorithm (SPA) and genetic algorithm (GA) to select an appropriate subset of wavenumbers for quadratic discriminant analysis (QDA). This novel approach envisions translation of this approach to ready use by practicing physicians and surgeons.

5

Material and Methods

Sample collection

Patients were identified from Upper Gastro-intestinal Multi-Disciplinary Team meetings and local hospital Pathology databases. Potential participants were identified prospectively and consented for biofluids (blood for plasma or serum, urine or saliva) and tissue was taken between October 2017 and June 2019 in a clinic or endoscopy setting. A power test (t-test-based with a 95% confidence level) was performed to determine the minimum sample size at 80% power, where a total of 82 samples was suggested. We collected 120 (plasma), 127 (saliva), 124 (serum) and 126 (urine) samples, which surpass the power of 80%. Biofluid specimens were categorised as follows: i) plasma: $n=35$ normal, $n=18$ inflammatory, $n=27$ Barrett's, $n=6$ LGD, $n=12$ HGD and $n=22$ OAC (set A); ii) saliva: $n=38$ normal, $n=19$ inflammatory, $n=27$ Barrett's, $n=6$ LGD, $n=12$ HGD and $n=25$ OAC (set B); iii) serum: $n=36$ normal, $n=19$ inflammatory, $n=28$ Barrett's, $n=6$ LGD, $n=12$ HGD and $n=23$ OAC (set C); iv) and, urine: $n=38$ normal, $n=19$ inflammatory, $n=27$ Barrett's, $n=6$ LGD, $n=11$ HGD and $n=25$ OAC (set D).

Ethical approval was granted by the East of England - Cambridge Central Research Ethics Committee from 2015 (Archival gastro-intestinal tissue, blood, saliva and urine collection; REC reference: 18/EE/0069; IRAS project ID: 242639). Ethics was also granted from the parent University (STEMH 909 application). All biofluids were stored in appropriate containers initially in a fridge at 4°C for up to 2 h. Plasma or serum samples were generated according to local protocols. Urine samples were centrifuged at 2200 rpm. All biofluids were then stored at -80°C.

Prior to slide preparation, biofluids samples were left to thaw in the fridge at 4°C. Thirty µL of individual biofluids (plasma, serum, saliva or urine) were pipetted onto naked FisherBrand™ slides for ATR-FTIR spectral analysis. Each slide was labelled with a specific GI (Gastrointestinal number) corresponding to its specific tissue pathological classification (*i.e.*, squamous tissue to adenocarcinoma). All slides were left to dry prior to transportation in wooden slide boxes to the laboratory for spectral analysis. All of the samples were stored in a de-humidified glass container to prevent condensation and physical damage.

ATR-FTIR spectroscopy

Spectroscopic interrogation of biofluid samples was performed at the Biomedical Research Laboratory at the University of Central Lancashire (UK). Histological diagnoses were unknown to those who performed IR spectroscopy. IR spectra were obtained using a Bruker TENSOR 27 FTIR spectrometer with Helios ATR attachment containing a diamond crystal (Bruker Optics Ltd, Coventry, UK) and operated using OPUS 6.5 software. Spectra were acquired from 10 independent sample locations. Data acquisition parameters were: 8 cm^{-1} spectral resolution giving 4 cm^{-1} data spacing, 32 scans, 6 mm aperture setting and 2× zero-filling factors. The ATR diamond crystal was washed with distilled water and dried with tissue paper between each sample and before each new slide. A background absorption spectrum (for atmospheric correction) was taken prior to each new sample.

Data analysis and chemometric methods

The data import, pre-treatment and construction of chemometric classification models were implemented in MATLAB R2014a software (MathWorks, USA) by using

PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., USA) and custom-made routines. Raw spectra were pre-processed by cutting between 1800 and 900 cm^{-1} (235 wavenumbers at 4 cm^{-1} spectral resolution), followed by rubberband baseline-correction and normalisation to the Amide I peak (*i.e.*, 1650 cm^{-1}). Before constructing the multivariate classification models (principal component analysis quadratic discriminant analysis, PCA-QDA; successive projections algorithm quadratic discriminant analysis, SPA-QDA; genetic algorithm quadratic discriminant analysis, GA-QDA) the samples were divided into training (60%), validation (20%) and prediction (20%) sets by the classic Kennard–Stone (KS)[37] uniform sampling algorithm applied to the IR spectra as shown in Table 1. The training samples were used in the modelling procedure, whereas the prediction set was only used in the final classification evaluation. The optimum number of variables for SPA-QDA and GA-QDA was determined according to an average risk G of misclassification. Such a cost function is calculated in the validation set as:

$$G = \frac{1}{N_V} \sum_{n=1}^{N_V} g_n \tag{1}$$

where g_n is defined as:

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m) \neq I(n)} r^2(x_n, m_{I(m)})} \tag{2}$$

and $I(n)$ is the index of the true class for the nth validation object x_n.

In this definition, the numerator is the squared Mahalanobis distance between object x_n (of class index $I(n)$) and the sample mean $m_{I(n)}$ of its true class. The denominator in Eq. (2) corresponds to the squared Mahalanobis distance between object x_n and the centre of the closest incorrect class. The minimum value of the cost function

8

(maximum fitness) is achieved when the selected variables from the original data are as close as possible to its true class and as distant as possible from its incorrect class in the validation set. The GA routine was carried out using 100 generations with 200 chromosomes each. Crossover and mutation probabilities were set to 60% and 1%, respectively. Moreover, the algorithm was repeated three times, starting from different random initial populations. The best solution (in terms of fitness value) resulting from three realizations of the GA was employed.

QDA classification score (Q_{ik}) is estimated using the variance-covariance matrix for each class k and an additional natural logarithm term, as follows:

$$Q_{ik} = (\mathbf{x}_i - \bar{\mathbf{x}}_k)^{\mathrm{T}} \mathbf{\Sigma}_k^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_k) + \log_e |\mathbf{\Sigma}_k| - 2 \log_e \pi_k \tag{3}$$

where $\mathbf{\Sigma}_k$ is the variance-covariance matrix of class k; and, $\log_e |\mathbf{\Sigma}_k|$ is the natural logarithm of the determinant of the variance-covariance matrix of class k. QDA forms a separated variance model for each class and does not assume that different classes have similar variance-covariance matrices, different to what is assumed by linear discriminant analysis (LDA).[38]

The calculation of figures of merit is a recommended standard practice to test model performance.[39] Herein, measures of test accuracy including sensitivity (portion of positive samples correctly classified), specificity (portion of negative samples correctly classified), and F-score, which is a general measurement of the model accuracy. These quality metrics are calculated using the following equations:

$$\text{Sensitivity (\%)} = \frac{\text{TP}}{\text{TP+FN}} \times \mathbf{100} \tag{3}$$

$$\text{Specificity (\%)} = \frac{\text{TN}}{\text{TN+FP}} \times \mathbf{100} \tag{4}$$

$$\text{F-score} = \frac{2 \times \text{SENS} \times \text{SPEC}}{\text{SENS+SPEC}} \tag{5}$$

where TP stands for true positives, TN for true negatives, FP for false positives and FN for false negatives. SENS stands for sensitivity and SPEC for specificity.

All selected wavenumbers derived from SPA-QDA and GA-QDA for oesophageal stages [*i.e.*, normal *vs.* inflammatory *vs.* Barrett's *vs.* low-grade dysplasia (LGD) *vs.* high-grade dysplasia (HGD) *vs.* oesophageal adenocarcinoma (OAC)] were confirmed by a two-tailed Student's *t*-test (95% confidence interval).

Results

The number of training, validation and prediction specimens (or spectra) in each biofluid category is summarised in Table 1.

Plasma dataset. Fig. 1A shows the average raw IR spectra derived from blood plasma for six oesophageal stages (Normal *vs.* Inflammatory *vs.* Barrett's *vs.* LGD *vs.* HGD *vs.* OAC), respectively. Overall, the IR spectra for oesophageal stages appear to overlap in the biochemical-cell fingerprint region (1800 cm^{-1} to 900 cm^{-1}), making it difficult to distinguish any subtle but significant differences. On closer analysis, notable distinguishing peaks that represent lipid functional groups could be observed around 1650 cm^{-1} (Amide I) and 1550 cm^{-1} (Amide II). In addition, peaks were observed in the region of 1050–1000 cm^{-1} (carbohydrates and collagen) and 1300 cm^{-1} to 1150 cm^{-1} (Amide III and $v_{as}PO_2^-$). The methyl groups of lipids and proteins major peaks could be found around 1400 cm^{-1}, 1260 cm^{-1} (Amide III), 1225 cm^{-1} ($v_{as}PO_2^-$) and 1080 cm^{-1} ($v_sPO_2^-$). To discriminate the six oesophageal stages, the spectral dataset was pre-processed using baseline correction and normalisation using the Amide I peak (Figure 1B). Average IR pre-processed spectra appear to overlap in the biochemical fingerprint

region (1800 cm^{-1} to 900 cm^{-1}) making spectral observation and the discovery of markers or signatures difficult. Therefore, chemometric techniques such as PCA-QDA, SPA-QDA and GA-QDA algorithms were adopted to classify Normal *vs.* Inflammatory *vs.* Barrett's *vs.* LGD *vs.* HGD *vs.* OAC based on their IR spectra.

A classification of the six oesophageal stages was developed by discriminant analysis using the IR spectra between 900 and 1800 cm^{-1}. Figure 1C shows the wavenumbers associated with class differences. The classification performance in the prediction set using GA-QDA was 100% accuracy, sensitivity, specificity and F-scores. Fig. 1D shows the predicted class achieved by the GA-QDA model for all classes based on 15 selected wavenumbers. An excellent classification of the samples was observed (only 3 errors in the training set and 1 error in the validation set). The PCA-QDA model using the scores on seven PCs (90% of the total data variance) achieved 100% for accuracy, sensitivity, specificity and F-scores for Normal, Inflammatory, Barrett's and OAC classes [see Supplementary Information (SI) Table S1]. SPA-QDA also achieved a very high accuracy for classification of LGD (100%) using 7 variables (namely, 1392, 1485, 1539, 1585, 1624, 1643, and 1681 cm^{-1}), as shown in Table S1 (see SI). Table S2 (see SI) lists the selected wavenumbers obtained by SPA-QDA and GA-QDA models for plasma samples with their tentative biomolecular assignments.

Saliva dataset. Figure 2A and 2B show the raw average IR spectra and average pre-processed (baseline correction and normalisation) in the biochemical-cell fingerprint region (1,800 cm^{-1} to 900 cm^{-1}) derived from the saliva for six oesophageal stages. There are notable differences in the wavenumber regions 1,000 cm^{-1} to 1,150 cm^{-1} (DNA/RNA region), 1,350 cm^{-1} to 1,500 cm^{-1} (Amide II protein region) and 1,530 cm^{-1} to 1,600 cm^{-1} (Amide I protein region) for all classes investigated. Although there is

some overlap amongst the average pre-processed spectra, the application of chemometric techniques (PCA-QDA, SPA-QDA or GA-QDA) exhibits good categorisation for all the classes (Normal *vs.* Inflammatory *vs.* Barrett's *vs.* LGD *vs.* HGD *vs.* OAC) (see SI Table S3).

As shown in Table S3 (see SI), classification of the six stages using saliva and the FTIR spectra between 900 and 1800 cm^{-1} using all models (PCA-QDA, SPA-QDA and GA-QDA) gives excellent classification accuracy (88.8–100%), though poor specificity in the LGD class using PCA-QDA and GA-QDA. SPA-QDA is the best classification model for saliva, with F-Score values ranging from 76.9 to 100%. The correct classification using PCA-QDA achieved 100% for all figures of merits for the inflammatory category in the prediction set. The accuracy and sensitivity were found to be >88% for all the other classes, using the scores on seven PCs from PCA (Figure 2C). The number of PCs is selected in Figure 2C based on the minimum number of PCs that generates the lowest power (eigenvalue), before the power follows a constant trend. Fig. 2D shows the predicted class achieved for PCA-QDA model for all classes. A good classification of the samples was observed (15 errors in the training set and 4 errors in the validation set). For SPA-QDA, 7 selected wavenumbers (902, 1014, 1099, 1589, 1643, 1697, and 1743 cm^{-1}) provided excellent classification, especially for the LGD class (100% predictive performance). The classification performance using GA-QDA was 100% for all figures of merit for the Barrett's class based on 14 selected wavenumbers (991, 1003, 1068, 1107, 1431, 1558, 1585, 1604, 1624, 1689, 1701, 1716, 1778 and 1786 cm^{-1}). Table S4 (see SI) lists the selected wavenumbers obtained by SPA-QDA and GA-QDA models for saliva samples with their tentative biomolecular assignments.

Serum dataset. Figures 3A and 3B show the average raw and pre-processed spectra for serum samples obtained from the six oesophageal stages. Classification of Normal *vs.* Inflammatory *vs.* Barrett's *vs.* LGD *vs.* HGD *vs.* OAC for serum was performed using discriminant analysis of the spectral bio-fingerprint region (1800-900 cm^{-1}). GA-QDA, using only 13 selected wavenumbers (1000, 1315, 1319, 1330, 1338, 1435, 1442, 1446, 1492, 1539, 1573, 1600 and 1654 cm^{-1}; Fig. 3C) achieved 100% for all figures of merit for all the OAC stage, as shown in Table S5 (see SI). The GA-QDA model demonstrated an excellent classification with 4 errors in the training set and 2 errors in the validation set (Figure 3D). For Normal *vs.* Inflammatory *vs.* Barrett's *vs.* OAC stages, the SPA-QDA model demonstrated sensitivities and specificities more than 71% (see SI Table S5) using only 7 variables (1041, 1477, 1539, 1593, 1631, 1662 and 1743 cm^{-1}). The PCA-QDA models for normal *vs.* Inflammatory *vs.* Barrett's *vs.* OAC stages produced sensitivity and specificity greater than 71% using seven PCA scores, which accounted for more than 90% of original data variance. Some notable differences can be observed in the wavenumber regions between 1,000 cm^{-1} to 1,338 cm^{-1} (DNA/RNA region), 1,435 cm^{-1} to 1,573 cm^{-1} (methyl groups of proteins and protein amide II absorption) and 1,600 cm^{-1} to 1,654 cm^{-1} (Amide I protein region) for all of the classes investigated. Table S6 (see SI) lists the selected wavenumbers obtained by SPA-QDA and GA-QDA models for serum samples with their tentative biomolecular assignments.

Urine dataset. Figure 4A and 4B show the raw and average pre-processed (baseline correction and normalization) spectral within the bio-fingerprint region (1,800 cm^{-1} to 900 cm^{-1}) derived from urine for all the six oesophageal stages. Although the discrimination for all classes on the basis of IR spectra is not straightforward due to the complexity of the spectra, good models using PCA-QDA, SPA-QDA and GA-QDA

were found for urine. The PCA-QDA model using urine spectra achieved 100% of classification for all figures of merit (accuracy, sensitivity, specificity and F-scores) for test samples using seven PCA scores as shown in Figure 4C. Fig. 4D demonstrated that the predicted classification performance achieved by the PCA-QDA model for all classes was 100%. As demonstrated in Table S7 (see SI), both SPA-QDA and GA-QDA models for urine samples achieved high sensitivity values for all classes (78.2–100%), but poor specificities particularly for Normal (SPA-QDA = 35.5%, GA-QDA = 62.5%), HGD (SPA-QDA = 33.3%) and OAC (SPA-QDA and GA-QDA = 20%).

The SPA-QDA model using only 7 selected wavenumbers (9096, 1242, 1577, 1600, 1651, 1681 and 1712 cm^{-1}) achieved also good classification, especially for Inflammatory, Barrett's and LGD classes. The predicted classification rate using GA-QDA demonstrates good results for Barrett's and HGD classes based on 19 selected wavenumbers (956, 995, 1018, 1030, 1095, 1118, 1141, 1253, 1334, 1381, 1431, 1446, 1500, 1550, 1562, 1681, 1724, 1735 and 1777 cm^{-1}). Spectral differences can be observed in the wavenumber regions between 956 cm^{-1} to 1,381 cm^{-1} (DNA/RNA region), 1,431 cm^{-1} to 1,562 cm^{-1} (Amide II region mainly stems from the C-N stretching and C-N-H bending vibrations) and 1,681 cm^{-1} to 1777 cm^{-1} (a high frequency vibration of an antiparallel β-sheet of Amide I and stretching vibration of lipids). Table S8 lists the selected wavenumbers obtained by SPA-QDA and GA-QDA models for urine samples with their tentative biomolecular assignments.

Discussion

There is no standard surveillance program for early detection of cancer in the oesophagus. The most reliable diagnostic measurement uses endoscopy with biopsy, which is advocated for screening oesophageal neoplasia especially in known high-risk

14

patients. Although endoscopy has been widely used in diagnosis and differential diagnosis of oesophageal diseases, approximately 20% of early oesophageal cancer lesions are barely visible to the naked eye. In order to have a robust surveillance program for early detection and improved cure rate for oesophageal adenocarcinoma, a skilled endoscopist and pathologist are necessary for proper mucosal sampling and histopathologic examination. However, the development of a quick, convenient, and inexpensive method for detecting early cancer or different stages such as Normal, Inflammatory, Barrett's, LGD, HGD and OAC can be useful specially to guide tissue biopsy, thus increasing the yield of dysplasia detection.

An approach to oesophageal cancer screening in the general population based on biofluids (blood plasma, serum, saliva and urine) interrogated by ATR-FTIR spectroscopy linked with feature selection methods for classification could be the potential to segregate stages of oesophageal adenocarcinoma. The use of ATR-FTIR spectroscopy coupled multivariate classification techniques (PCA-QDA, SPA-QDA and GA-QDA) in identifying oesophageal stages of disease to adenocarcinoma has achieved excellent accuracy, sensitivity and specificity, encouraging investigation of screening for others cancers with known markers.

ATR-FTIR spectroscopy was employed to predict six oesophageal stages in four different biofluids (plasma, saliva, serum and urine). PCA-QDA and GA-QDA models were found to give the best class differentiation compared to the SPA-QDA. GA-QDA successfully detected the biochemical alterations in the oesophageal stages based on plasma (100% accuracy, sensitivity, specificity and F-scores) using a few wavenumbers or biomarkers. For this model, several selected wavenumbers appear to be of particular interest, especially at 999 cm^{-1} and 1381 cm^{-1}, representing the ring stretching vibrations mixed strongly with CH in-plane bending and C-O stretching, respectively. Others

selected wavenumbers found by GA-QDA using plasma can be found in Table S2. In general, the spectral alterations responsible for discrimination of oesophageal stages based on plasma blood were mainly associated with DNA/RNA and proteins at wavenumbers between 929 cm^{-1} and 1,431 cm^{-1}. The major advantage for this model, using blood plasma, is the minimal sample preparation for FTIR analysis.

Since a direct, easy-to-use, compact test device for oesophageal stages in human saliva is currently not available, we suggest ATR-FTIR spectroscopy combined with multivariate classification techniques for the development of a direct test that meets these challenges. Although saliva has a complex biology (components are produced mainly by acinar cells and are delivered to the oral cavity by a cell-lined duct system), we believe that saliva analysis based on ATR-FTIR is a powerful diagnostic tool in cancer. As demonstrated by the PCA-QDA model, the oesophageal stages diagnostic based on the saliva spectral data was found to have good accuracy, sensitivity, specificity and F-scores values for all classes, especially for the Inflammatory oesophageal stage. The selected wavenumbers found by SPA-QDA and GA-QDA models using saliva samples can be found in Table S2 (see SI). Several selected wavenumbers appear to be of particular interest, namely, the variables at 1604, 1624 and 1643 cm^{-1}, representing the adenine vibration in DNA, peak of nucleic acids due to the base carbonyl stretching, and ring breathing mode and amide I band (arises from C=O stretching vibrations), respectively.

The classification rates for plasma samples were found to be better than those of saliva and serum for all classes. One explanation for such a difference is that plasma contains thousands of biomolecules at various concentrations while serum and saliva

may contain a more limited number of analytes resulting in fewer variations in peak intensities or shifts of the IR spectrum.

Problems can arise with liquid samples, such as urine, because very low concentration components may not be detected. Low concentrations can be overcome by drying samples directly on to the crystal to increase their concentrations and, therefore, increase their signal intensities. We believe that clinically relevant levels could be detected in ATR-FTIR spectra of dried insoluble fractions of urine samples without any requirement for chemical manipulation. Urine can be collected non-invasively and without the need for a trained professional to be present. It can also be collected frequently and stored for several days. It is therefore ideal as a diagnostic medium if it can provide clinically useful information. Urine samples for this present study demonstrated excellent discrimination for all stages using urine samples based on PCA-QDA models. SPA-QDA and GA-QDA presented relatively well segregation amongst all the classes using a few selected wavenumbers, as can seen in Table S8 (see SI). Despite relatively high values of accuracies and sensitivities, these two algorithms, in particular SPA-QDA, produced low specificity values (<50%) for classes Normal, HGD, and OAC (Table S7). The specificity is a measure of the classification fail towards one class. For example, a specificity of 100% in OAC means that no misclassification in the other 5 classes was assigned as OAC (there is no false positive for OAC); while a specificity of 20% in OAC means that most samples wrongly classified in this model were assigned as OAC. This affects the model classification performance, despite the satisfactory values of accuracy and sensitivity in general. Thus, for urine samples, PCA-QDA is the most reliable model.

The classification performance of the spectroscopy-based models to determine oesophageal stage are a result of a combination between the algorithm being used and the nature of the sample measured. Although the type of classifier employed in this study is the same for all algorithms (QDA), the feature extraction (PCA) and feature selection (SPA and GA) methods work in different manner, thus leading the models to different results. PCA reduces the spectral dataset to features representing the main sources of variance in the data, but not necessarily these sources of variation are correlated with differences between the samples; and, SPA and GA, which are iterative algorithms of variable selection, work by reducing data collinearity (SPA) and by mimicking the process of natural selection in a computation fashion (GA).[40] Differently from PCA, both SPA and GA act on the original sample space, while PCA projects the sample on a orthogonal space; and SPA and GA are more complex algorithms, where more parameters need to be optimized and the computational-cost is higher. Therefore, choosing the right algorithm for data analysis is an empirical process.

Spectroscopy of urine has been already used to diagnostic several cancers, such as endometrial and ovarian cancer, where different spectral markers were found to be associated with the pathologic stage.[41] Molecular signatures of oesophageal cancer and Barrett's oesophagus have been found in urine;[42] and other biofluids, such as blood[43] and saliva,[44] are known to contain key molecular signatures related to oesophageal cancer. For these reasons, and confirmed by the results reported herein, we believe that IR spectroscopy can be used to discriminate oesophageal transformation to adenocarcinoma based on these biofluids.

Conclusion

The results of this study show that ATR-FTIR spectroscopy coupled with multivariate classification algorithms (PCA-QDA, SPA-QDA and GA-QDA), result in a powerful alternative approach for detection of oesophageal stages of disease to OAC. Herein, we present a new, rational and convenient approach to different biofluids (plasma, saliva, serum and urine) using ATR-FTIR spectroscopy, opening a new level of non-invasive diagnostic tool in this field. We also demonstrate a fast, clean, and non-destructive methodology involving minimal sample preparation to categorise the samples. For urine samples, the resulting PCA-QDA model successfully detect biochemical alterations at the maximum classification rate (100%) for different figures of merit (accuracy, sensitivity, and F-scores). This method makes it possible to detect all the oesophageal stages to adenocarcinoma without special sample preparation and reagents, from a minimal sample volume and (almost) immediately after sample collection. In this pilot study, we have demonstrated for the first time that saliva- and urine-based ATR-FTIR spectroscopy coupled with a multivariate classification algorithm has the potential to discriminate oesophageal stages. Further work with biofluids and spectroscopic analysis should be performed in the future to validate these encouraging results.

Conflicts of Interest

There are no conflicts to declare.

Acknowledgements

References

1   H. Pohl and H. G. Welch, *J. Natl. Cancer Inst.*, 2005, **97**, 142–146.

2   L. H. Moyes and J. J. Going, *J. Clin. Pathol.*, 2011, **64**, 742–750.

3   A. Wong and R. C. Fitzgerald, *Clin. Gastroenterol. Hepatol.*, 2005, **3**, 1–10.

4   S. Bhat, H. G. Coleman, F. Yousef, B. T. Johnston, D. T. McManus, A. T. Gavin and L. J. Murray, *J. Natl. Cancer Inst.*, 2011, **103**, 1049–1057.

5   S. J. Spechler, P. Sharma, R. F. Souza, J. M. Inadomi and N. J. Shaheen, *Gastroenterology*, 2011, **140**, e18–e13.

6   A. Bhardwaj, D. B. Stairs, H. Mani and T. J. McGarrity, *Patholog. Res. Int.*, 2012, **2012**, 1–20.

7   S. J. Spechler. *JAMA,* 2013, **310**, 627–636.

8   C. Amadi and P. Gatenby, *World J Gastroenterol.*, 2017, **23**, 5051–5067.

9   R. Krishnamoorthi, S. Singh, K. Ragunathan, K. Visrodia, K. K. Wang, D .A. Katzka and P. G. Iyer, *Clin. Gastroenterol. Hepatol.*, 2018, **16**, 1046–1055.

10   R. Milind and S. E. Attwood, *World J Gastroenterol.* 2012, **18**, 3483–91.

11   M. A. Kara, F. P. Peters, W. D. Rosmolen, K. K. Krishnadath, F. J. ten Kate, P. Fockens and J. J. G. H. Bergman, *Endoscopy*, 2005, **37**, 929–936.

12   J. Pohl, O. Pech, A. May, H. Manner and A. Fissler-Eckhoff, *Am J Gastroenterol*, 2010, **105**, 2350–2356.

13   P. J. Trivedi and B. Braden, *QJM*, 2013, **106**, 117–131.

14   J. Mannath, V. Subramanian, C. J. Hawkey and K. Ragunath, *Endoscopy*, 2010, **42**, 351–359.

15   R. S. DaCosta, B. C. Wilson and N. E. Marcon, *Dig. Endosc.*, 2003, **15**, 153–173.

16   M. A. Kara, R. S. DaCosta, C. J. Streutker, N. E. Marcon, J. J. G. H. M. Bergman and B. C. Wilson, *Dis Esophagus*, 2007, **20**, 141–150.

17   W. L. Curvers, F. G. van Vilsteren, L. C. Baak, C. Böhmer, R. C. Mallant-Hent, A. H. Naber, A.van Oijen, C. Y. Ponsioen, P. Scholten, E. Schenk, E. Schoon, C. A. Seldenrijk, G. A. Meijer, F. J. ten Kate and J. J. Bergman, *Gastrointest Endosc*, 2011, **73**, 195–203.

18   K. Egger, M. Werner, A. Meining, R. Ott, H. D. Allescher, H. Hoffler, M. Classen and T. Rosch, *Gut*, 2003, **52**, 18–23.

19   F. L. Martin, J. G. Kelly, V. Llabjani, P. L. Martin-Hirsch, I. I.Patel, J. Trevisan, N. J. Fullwood and M. J. Walsh, *Nat. Protoc.*, 2010, **5**, 1748–1760.

20   M. J. Baker, J. Trevisan, P, Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S.W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J.Walsh, B. R.Wood, P. Gardner and F. L. Martin, *Nat. Protoc.*, 2014, **9**, 1771–1791.

21   S. De Bruyne, M. M. Speeckaert and J. R. Delanghe, *Crit. Rev. Clin. Lab. Sci.*, 2018, **55**, 1–20.

22   G. Theophilou, K. M. G. Lima, M. Briggs, P. L.Martin-Hirsch, H. F. Stringfellow and F. L. Martin, *Analyst*, 2016, **141**, 585–94.

23   N. C. Purandare, J. Trevisan, I. I. Patel, K. Gajjar, A. L. Mitchell, G. Theophilou, G. Valasoulis, M. Martin, G. von Bünau, M. Kyrgiou, E. Paraskevaidis, P. L. Martin-Hirsch, W. J. Prendiville and F. L. Martin, *Bioanalysis*, 2013, **5**, 2697–711.

24   J. G. Kelly, A. A. Ahmadzai, P. Hermansen, M. A. Pitt, Z. Saidan, P. L. Martin-Hirsch and F. L. Martin, *Anal. Bioanal. Chem.*, 2011, **401**, 957–967.

25   J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D. Jenkinson, R. W. Lea, C. Walker and M. J. Baker, *J. Biophotonics*, 2014, **7**, 189–199.

26   G. Theophilou, K. M. G. Lima, M. Briggs, P. L. Martin-Hirsch, H. F. Stringfellow and F.L. Martin, *Sci. Rep,* 2015, **5**, 13465.

27   C. Petibois, B. Drogat, A. Bikfalvi, G. Déléris and M. Moenner, *FEBS Lett.*, 2007, **581**, 5469–5474.

28   F. Andleed, Hafeezullah, A. Atiq, M. Atiq and S. Malik, *Cancer Biomarkers*, 2018, **1**, 1–8.

29   M. Wu, W. Zhang, P. Tian, X. Ling and Z. Xu, *Int J Clin Exp Med*, 2016, **9**, 2351–2358.

30   Q. Li, X. Sun, Y. Xu, L. Yang, Y. Zhang, S. Weng, J. Shi and J. Wu, *World J Gastroenterol.*2005, **11**, 3842–3845.

31   Q. Li, C. Hao, X. Kang, J. Zhang, X. Sun, W. Wang and H. Zeng, *Sensors*, 2017, **17**, 1–9.

32   G. J. Vazquez-Zapien, M. M. Mata-Miranda, V. Sanchez-Monroy, R. J. Delgado-Macuil, D. G. Perez-Ishiwara and M. Rojas-Lopez, *Stem Cells Int.*, 2016, **2016**, 1–10.

33   D. E. Mazia., M. T. Do, F. M. Shamji, S. R. Sundaresan, D. G. Perkins and P. T.

Wong, *Cancer Detect. Prev.*, 2007, **31**, 244–253.

34   L. Quaroni and A. G. Casson, *Analyst*, 2009, **134**, 1240–1246.

35   J. S. Wang, Y. Z. Xu, J. S. Shi, L. Zhang, X. Y. Duan, L. M. Yang, Y. L. Su, S. F. Weng, D. F. Xu and J. G. Wu, *Guangpuxue Yu Guangpu Fenxi.* 2003, **23**, 863–865.

36   T. D. Wang, G. Triadafilopoulos, J. M. Crawford, L. R. Dixon, T. Bhandari, P. Sahbaie, S. Friedland, R. Soetikno and C. H. Contag, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 15864–15869.

37   R. Kennard and L. Stone, *Technometric*, 1969, **11**, 137–148.

38   S. J. Dixon and R. G. Brereton, *Chemom. Intell. Lab. Syst.*, 2009, **95**, 1–17.

39   K. T. Cheung, J. Trevisan, J. G. Kelly, K. M. Ashton, H. F. Stringfellow, S. E. Taylor, M. N. Singh, P. L. Martin-Hirsch and F. L. Martin, *Analyst*, 2011, **136**, 2047.

40   C. L. M. Morais, M. Paraskevaidi, L. Cui, N. J. Fullwood, M. Isabelle, K. M. G. Lima, P. L. Martin-Hirsch, H. Sreedhar, J. Trevisan, M. J. Walsh, D. Zhang, Y. G. Zhu and F. L. Martin, *Nat. Protoc.*, 2019, 14, 1546–1577.

41   M. Paraskevaidi., C. L. M. Morais, K. M. G. Lima, K. M. Ashton, H. F. Stringfellow, P. L. Martin-Hirsch and F. L. Martin, *Analyst*, 2018, **143**, 3156-3163.

42   V. W. Davis, D. E. Schiller, D. Eurich and M. B. Sawyer, *World J. Surg. Oncol.*, 2012, **10**, 271.

43   A. T. Kunzmann, Ú. C. McMenamin, A. D. Spence, R. T. Gray, L. J. Murray, R. C. Turkington and H. G. Coleman, *Eur. J. Gastroenterol. Hepatol.*, 2018, **30**, 263–273.

44   O. Rapado-González, B. Majem, L. Muinelo-Romay, R. López-López and M. M. Suarez-Cunqueiro, *Int. J. Mol Sci.*, 2016, **17**, 1531.

Figures Captions

Figure 1: Comparison of Normal/Inflammatory/Barrett's/LGD/HGD/OAC oesophageal stages using plasma samples. The panel shows: (A) Average raw in mid-IR region of 1800 cm^{-1} to 900 cm^{-1} and (B) average pre-processed IR spectra obtained from all stages segregated into Normal (black colour) *vs.* Inflammatory (blue colour) *vs.* Barrett's (green colour) *vs.* LGD (yellow colour) *vs.* HGD (magenta colour) *vs.* OAC (red colour). (C) 15 selected wavenumbers by GA-QDA. (D) Predicted class *vs.* samples used for training and prediction sets (dashed box), where each circle marker represents one sample.

Figure 2: Comparison of Normal/Inflammatory/Barrett's/LGD/HGD/OAC oesophageal stages using saliva samples. The panel shows: (A) Average raw in mid-IR region of 1800 cm^{-1} to 900 cm^{-1} and (B) average pre-processed IR spectra obtained from all stages segregated into Normal (black colour) *vs.* Inflammatory (blue colour) *vs.* Barrett's (green colour) *vs.* LGD (yellow colour) *vs.* HGD (magenta colour) *vs.* OAC (red colour). (C) Singular value decomposition (SVD) *vs.* Number of principal component (PC) obtained by PCA-QDA, where Power represents the eigenvalue. (D) Predicted class *vs.* samples used for training and prediction sets (dashed box), where each circle marker represents one sample.

Figure 3: Comparison of Normal/Inflammatory/Barrett's/LGD/HGD/OAC oesophageal stages using serum samples. The panel shows: (A) Average raw in mid-IR region of 1800 cm^{-1} to 900 cm^{-1} and (B) average pre-processed IR spectra obtained from all stages segregated into Normal (black colour) *vs.* Inflammatory (blue colour) *vs.* Barrett's (green colour) *vs.* LGD (yellow colour) *vs.* HGD (magenta colour) *vs.* OAC (red colour). (C) 13 selected wavenumbers by GA-QDA. (D) Predicted class *vs.* samples

used for training and prediction sets (dashed box), where each circle marker represents one sample.

Figure 4: Comparison of Normal/Inflammatory/Barrett's/LGD/HGD/OAC oesophageal stages using urine samples. The panel shows: (A) Average raw in mid-IR region of 1800 cm^{-1} to 900 cm^{-1} and (B) average pre-processed IR spectra obtained from all stages segregated into Normal (black colour) *vs.* Inflammatory (blue colour) *vs.* Barrett's (green colour) *vs.* LGD (yellow colour) *vs.* HGD (magenta colour) *vs.* OAC (red colour). (C) Singular value decomposition (SVD) vs. Number of principal component (PC) obtained by PCA-QDA, where Power represents the eigenvalue. (D) Predicted class plot *vs.* samples used for training and prediction sets (dashed box), where each circle marker represents one sample.

Tables Legends

Table 1: Number of training, validation and prediction specimens in each biofluids category.

Category – biofluids	Training	Validation	Prediction
Normal – plasma	21	7	7
Normal – serum	22	7	7
Normal – saliva	22	8	8
Normal – urine	22	8	8
Inflammatory – plasma	10	4	4
Inflammatory – serum	11	4	4
Inflammatory – saliva	11	4	4
Inflammatory – urine	11	4	4
Barrett's – plasma	17	5	5
Barrett's – serum	18	5	5
Barrett's – saliva	17	5	5
Barrett's – urine	17	5	5
LGD – plasma	3	1	2
LGD – serum	3	1	2
LGD – saliva	3	1	2
LGD – urine	3	1	2
HGD – plasma	7	2	3
HGD – serum	7	2	3
HGD – saliva	7	2	3
HGD – urine	6	2	3
OAC – plasma	12	5	5
OAC – serum	14	4	4
OAC – saliva	15	5	5
OAC – urine	15	5	5

Figure 1
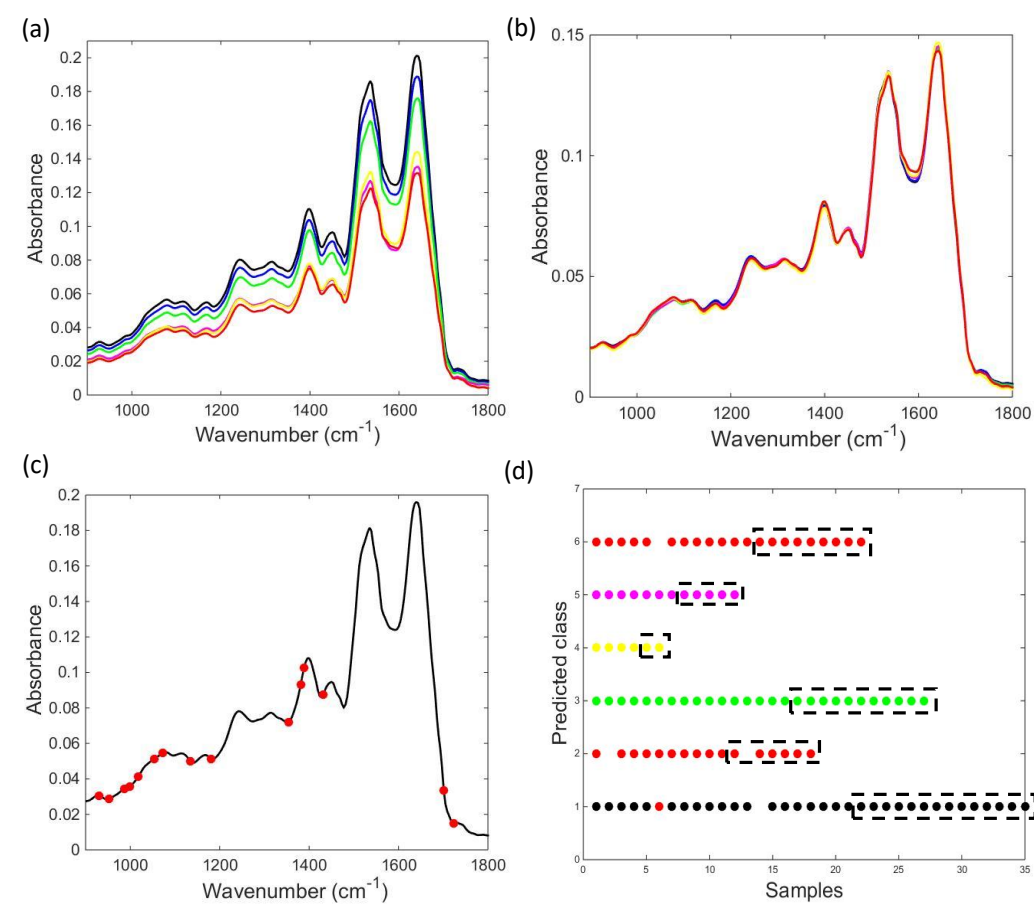
Figure 2

Figure 3

Figure 4