scientific reports

Check for updates

OPEN

On-chip Raman spectroscopy of live single cells for the staging of oesophageal adenocarcinoma progression

Alisha Farooq[1,2], Christopher D. Wood[3], John E. Ladbury[2] & Stephen D. Evans[1✉]

The absence of early diagnosis contributes to oesophageal cancer being the sixth most common cause of global cancer-associated deaths, with a 5-year survival rate of < 20%. Barrett's oesophagus is the main pre-cancerous condition to adenocarcinoma development, characterised by the morphological transition of oesophageal squamous epithelium to metaplastic columnar epithelium. Early tracking and treatment of oesophageal adenocarcinoma could dramatically improve with diagnosis and monitoring of patients with Barrett's Oesophagus. Current diagnostic methods involve invasive techniques such as endoscopies and, with only a few identified biomarkers of disease progression, the detection of oesophageal adenocarcinoma is costly and challenging. In this work, single-cell Raman spectroscopy was combined with microfluidic techniques to characterise the development of oesophageal adenocarcinoma through the progression of healthy epithelial, Barrett's oesophagus and oesophageal adenocarcinoma cell lines. Principal component analysis and linear discriminant analysis were used to classify the different stages of cancer progression. with the ability to differentiate between healthy and cancerous cells with an accuracy of 97%. Whilst the approach could also separate the dysplastic stages from healthy or cancer with high accuracy—the intra-class separation was approximately 68%. Overall, these results highlight the potential for rapid and reliable diagnostic/ prognostic screening of Barrett's Oesophagus patients.

Cases of benign oesophageal diseases such as gastro-oesophageal reflux disease (GORD), eosinophilic oesophagitis, and Barrett's oesophagus are increasing in prevalence, with the latter classified as the main pre-cancerous condition for oesophageal adenocarcinoma (OAC)[1]. Characterised by the replacement of squamous oesophageal epithelia by a mosaic of intestinal and gastric cell types, Barrett's Oesophagus is an archetypal metaplastic condition found in up to 10% of GORD patients[2]. Barrett's Oesophagus occurs as an adaptive response to recurrent acid and bile exposure of the epithelial squamous mucosa, leading to DNA damage, with 0.12–0.6% of patients annually developing low-grade and high-grade dysplasia leading to invasive OAC[3]. Risk factors such as obesity and acid reflux can have a synergistic effect due to increased intra-abdominal pressure and obesity-related metabolic syndrome. Other contributing factors include being of the male sex (7:1 to females), low vegetable and fruit intake, and high red meat, alcohol and tobacco intake[2]. Early detection of dysplastic changes within Barrett's Oesophagus would allow intervention prior to cancer development. Currently, there is no treatment leading to complete recovery once a patient has been diagnosed.

The integration of multiple disciplines, such as pathology, physiology, and analytical chemistry, has led to the increasing use of microfluidic platforms in cancer research over the last three decades[4,5]. The customisable nature of microfluidic chips has allowed the development of platforms for studying systems from single-cell to patient-derived tissues[6]. The high-throughput, low cost and improved efficiency of these devices provide the ideal platform for identification and characterisation of cancer cell types and metastatic patterns[7,8]. Microfluidic chips have been utilised with other methods of recent cancer research in studies including: tumour cell detection in blood[9,10], angiogenesis within 3D culture[11], metastasis of breast cancer[12,13], and biochemical analysis using Raman spectroscopy, with the latter combination becoming increasingly used[14–17].

Raman spectroscopy has been previously used for the detection and diagnosis of cancers[18–20], including to characterise cells during surgical endoscopy[21,22]. Confocal Raman spectroscopy offers diffraction-limited spatial

[1]School of Physics and Astronomy, University of Leeds, Leeds, UK. [2]School of Molecular and Cellular Biology, University of Leeds, Leeds, UK. [3]School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK. ✉email: S.D.Evans@leeds.ac.uk

resolution (< 1 μm), which allows micro-spectroscopic analysis of femtolitre volumes within single cells[23]. For Raman spectroscopy to be utilised for single-cell microfluidic clinical screening, an essential requirement is the ability to trap individual cells from a suspension within a flow system[24]. An emerging method for identifying markers across oesophageal tissues[1,25–27], Raman spectroscopy has potential use as a diagnostic tool for the evaluation of both cancerous and pre-cancerous lesions both in and ex vivo. This includes important studies identifying spectral features associated with different stages of oesophageal adenocarcinoma development using both patient tissues and biofluids[25,26].

Whilst previous publications detail the use of Raman probes to characterise cells during surgical endoscopy[21,22], there are no reports characterising the biochemical progression between each stage of Barrett's dysplasia at a single cell level. Herein, we report live single-cell spectral analysis of six representative cell lines ranging from a healthy oesophageal epithelium (HET-1A) model to oesophageal adenocarcinoma stage III gastric cardia cells (OE19). The four progressive phases of Barrett's oesophagus, between the healthy epithelium (HET-1A) and the adenocarcinoma (OE19), are represented by cell lines characteristic of low-grade dysplasia (CP-A), mild grade dysplasia (CP-B), moderate grade dysplasia (CP-C) and severe grade dysplasia (CP-D). The microfluidic platform was designed to trap single cells, allowing time-resolved Raman studies to be performed. Principal component analysis (PCA) and linear discriminant analysis (LDA) were used to identify spectral differences associated with progression throughout each stage of diease development with average classification accuracies for all results over 75%.

Results

Single cell trapping

Cells were trapped in the microfluidic device at a flow rate of 3 μL/min, and subsequently maintained at 37.5 °C, with a constant flow of fresh culture media (1 μL/min) for the duration of the experiment. The polydimethylsiloxane (PDMS) chip consisted of an inlet, an outlet, and a chamber containing ~ 1200 cell traps. The traps were designed to minimise the cell-PDMS interaction and PDMS contributions to the Raman spectra. Post-Raman spectral acquisition, the on-chip cell viability was monitored over five hours using a Calcein AM live stain and a Sytox Red dead stain. Figure 1 shows an overview of the chip and trap design, together with bright-field and fluorescence (Calcein AM) images of trapped CP-A cells. Previously reported single-cell hydrodynamic traps have exhibited trapping efficiency between 50 and 80% for different design variations[28,29]. The trap design used here (Fig. 1) exhibited a trapping efficiency of 74% (± 6%), averaged over more than 15 independent experiments.

Changes with disease development

Raman spectra were obtained from trapped live single cells across the cancerous cell line (OE19), those representing the healthy epithelium (HET-1A) and the intermediary Barrett's oesophagus (CP-A to CP-D). Figure 2 shows the averaged spectrum and standard error for each cell line, averaging over 100 single cells per sample type. All bands were preprocessed using EMSC and normalised to the Amide I peak at 1656 cm^{-1}. Increased aromatic amino acids (tryptophan[30,31], tyrosine[23,32–34] and phenylalainine[32,34]) have been associated with cancers such as breast[35] and colorectal[36], as well as mutations associated with the acquisition of transforming functions in cancer development[37]. Additional bands related to increased aromatic acids are present within the OE19 cell line, such as the tryptophan shoulder at 1009 cm^{-1}[35,37].

Nucleic acid and protein-related bands, such as Amide III and Amide I, are labelled and present across all six cell lines. DNA-associated bands, as seen at 1089 cm^{-1}, are common within transformed cancerous cells[34], with an increased DNA concentration associated with disease progression. An increase in protein signature is
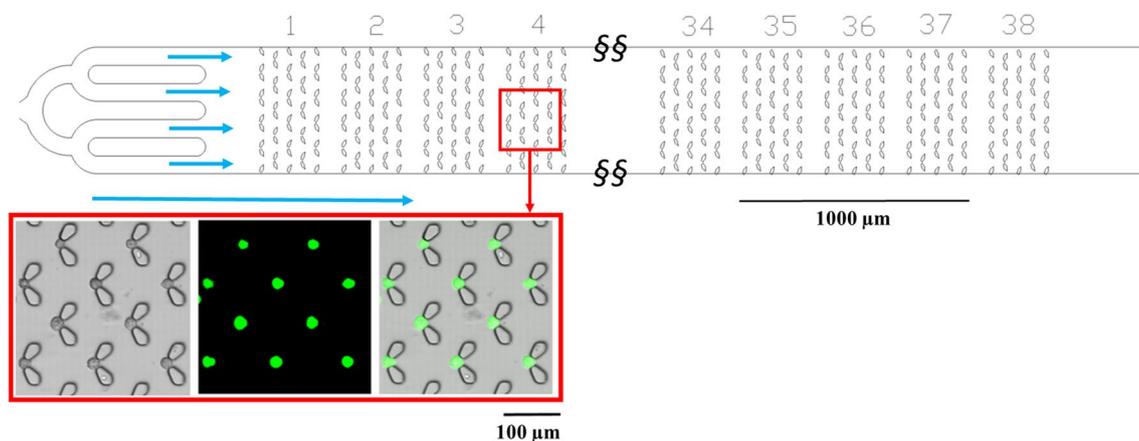

Figure 1. Top: Schematic design of the microfluidic cell traps device. Fluid flow direction is indicated by the blue arrows. Scale bar 1000 μm. Bottom: example of CP-A cells trapped in the device and stained using Calcein AM (live) and Sytox Red (dead) to determine cell viability over 5 h. Bottom left: bright-field image, bottom middle: fluorescence and bottom right: bright field and fluorescence overlay.
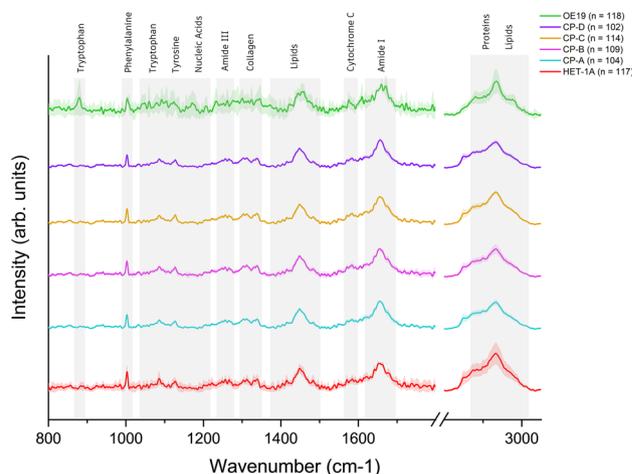
Figure 2. Single-cell Raman spectra representing the progression from healthy squamous oesophageal epithelia (HET-1A: red) through non-dysplastic Barrett's (CP-A: cyan), mild dysplastic Barrett's (CP-B: pink), moderate dysplastic Barrett's (CP-C: orange) and severe dysplastic Barrett's (CP-D: purple), to oesophageal adenocarcinoma (OE19: green) cells. All spectra were normalized to the Amide I band at 1656 cm^{-1}. The shaded area around each cell line is representative of the standard deviation. The grey shading across all cell lines highlights bands of interest (labelled). n = number of single cell Raman spectra. The high wavenumber region (CH$_2$CH$_3$) has been multiplied by a factor of 0.25 to fit on the same scale.

observed via the increase in the CH$_3$ stretching (2881 cm^{-1} and 2980 cm^{-1}) for the OE19 (cancer) cells. Such an increase in protein content is consistent with previously reported data[25,38,39].

The CH$_3$ and CH$_2$ twisting modes previously attributed to collagen[40] were observed between 1310 and 1340 cm^{-1}[37,42]. Lipid-associated bands are reported as being attributed to the acyl (C–C) backbone in lipids[41,42], CH$_3$ rocking/C–O vibrations[26], bending modes of C–H in CH$_2$ moieties[43,44] as well as methylene deformation[34]. These changes in cellular lipid quantity are significant biomarkers of abnormal membrane composition, characteristic of neoplastic cells[45].

Comparison of healthy versus cancer cells

Analysis was carried out to determine if the OE19 oesophageal adenocarcinoma cell line could be distinguished from HET-1A (healthy) cells. PCA was conducted as an unsupervised technique with the goal of maximising and identifying the variance of the data along the principal components (PCs). These PCs can be visualised and plotted to identify features within the spectra that help distinguish the different cell lines. PCA was undertaken using the first 15 PCs of a truncated Raman spectra (Fig. 3a). Differences across the first 5 PCs, accounting for > 80% variance, are shown in Fig. 3b, with bands showing significant differences highlighted in blue. The cumulative variance and variance explained for all 15 PCs is shown in Fig. 3c. Variation within the amino acid bands (tryptophan, tyrosine and cytosine) are seen in the first 5 PCs at 880 cm^{-1}, 1171 cm^{-1} and 1610 cm^{-1} respectively. Additional bands showing variation across all 5 PCs include DNA (1089 cm^{-1}), collagen (1310–1340 cm^{-1}), methylene deformation (1450 cm^{-1}) and Amide I (1656 cm^{-1}).

Further analysis was undertaken using LDA, a supervised technique used for classification, by maximising the separation between different classes within a dataset via their discriminant functions. The LDA was trained on 75% of the data and tested on the remaining 25% using a fourfold cross-validation approach. The combined PCA-LDA was able to separate the datasets into distinct classes with high accuracy. The histogram (Fig. 3d) indicates a clear separation of both populations. It is also evident that the representative healthy cell line (HET-1A, red) provides a narrower distribution compared to the cancerous equivalent (OE19, green).

The confusion matrix obtained from this analysis is shown in Fig. 3e. Whilst all HET-1A cells were correctly categorised, approximately 6% of OE19 cells were mistaken for HET-1A, this is presumably a consequence of their heterogeneity. The precision, recall and F1-score (defined in Supplementary Fig. S1) were determined to be 1, 0.972 and 0.942 respectively. The average accuracy for the classification of healthy vs. cancer single cells was 97%.

Analysis of Barrett's oesophagus cell lines

Intermediate between the healthy (HET-1A) and cancerous (OE19) stages are the progressive stages of Barrett's oesophagus, represented by the cell lines CP-A (non-dysplastic) and CP-D (severe dysplastia). Shown in Supplementary Fig. S2, the spectra for each cell line was truncated to 800–1800 cm^{-1} prior to PCA-LDA. Analysis of the first 5 PC loadings, covering 85% variance (Fig. 4b), confirms the influence of amino acid bands, with phenylalanine and tryptophan bands present across all 5 PCs. Other cellular components such as Amide III, Lipids and Amide are present in PCs 1 and 5, PCs 2, 4 and 5, and PCs 2 and 5, respectively. Additionally, between disease stages, the Cytochrome C (Cyt C) band (1583 cm^{-1}) is present within PC1 and PC2. This peak can be attributed to the presence of aromatic side chains within the Cyt C compound[46–48].
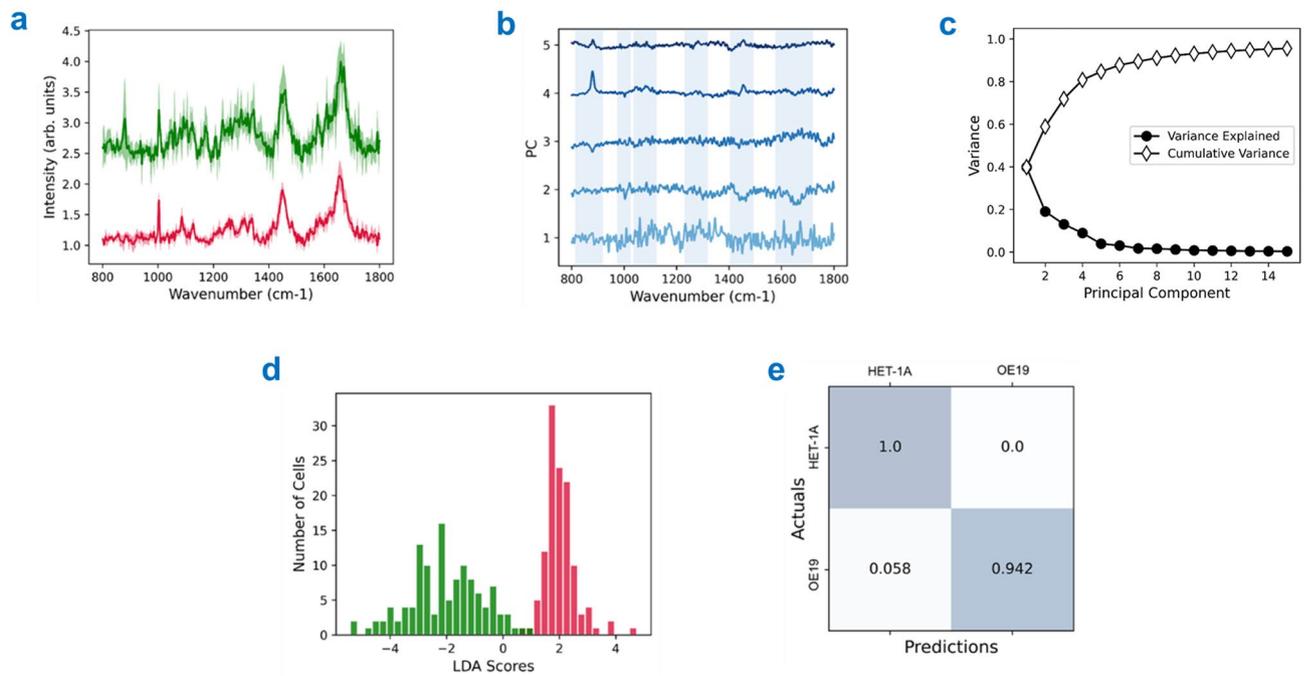
Figure 3. PCA-LDA of HET-1A (red) and OE19 (green) EMSC-corrected spectral data. (**a**) Truncated Raman datasets between 800 and 1800 cm⁻¹ used for analysis. (**b**) Principal component loadings taken from PCA. Scores 1–5 are shown, accounting for > 85% variability. Bands of interest are highlighted in blue. (**c**) Cumulative variance and variance explained for the first 15 PCs. (**d**) LDA histogram displaying the spread of the scores for each dataset. (**e**) Confusion matrix for statistical analysis of the PCA-LDA. The outputted values are representative of the False Positive (0.058), True Positive (1), False Negative (0.942) and True Negative (0).
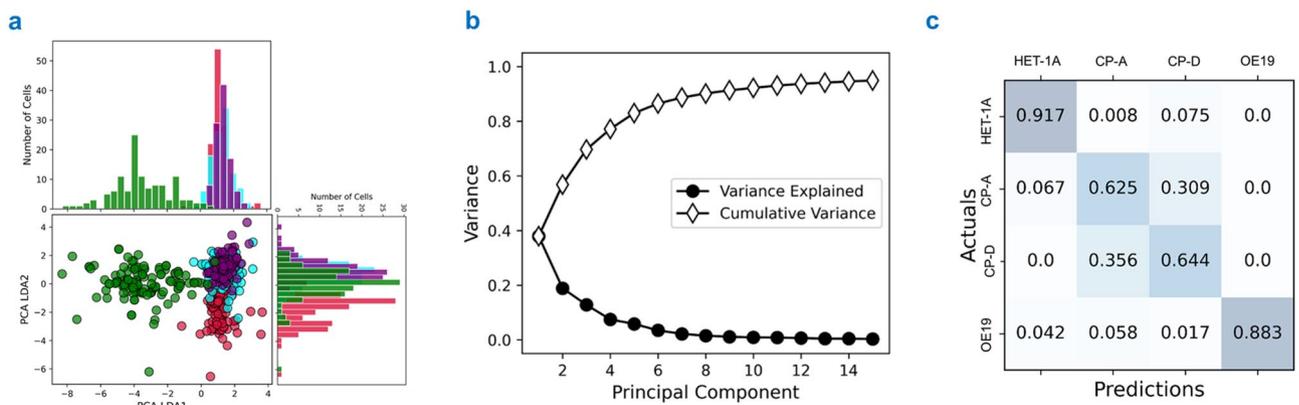

Figure 4. (**a**) 2D plotted PCA LDA1 vs PCA LDA2 scores for HET-1A (red), CP-A (cyan), CP-D (purple) and OE19 (green) and their relative LDA histograms. (**b**) Cumulative variance and variance explained for the first 15 PCs. (**c**) Confusion matrix for statistical analysis of the PCA-LDA.

The 2D scatter plot for the LDA is displayed in Fig. 4a, with HET-1A (red), CP-A (cyan), CP-D (purple), and OE19 (green), along with their respective histograms.

From this scatter plot we see that the OE19 dataset is separated from the others and also shows greater scatter as might be expected from a cancerous dataset with uncontrolled mutation. The HET-1A appears to be separated from CP-A and CP-D datasets in LDA2, whilst the CP-A and CP-D cells appear to exhibit strong overlap. This is further supported by looking at the confusion matrix in Fig. 4c, which suggests that neither of the Barrett's Oesophagus cell lines (CP-A and CP-D) nor the healthy cell line (HET-1A) were mis-categorised as the cancer cell line (OE19). The HET-1A cells were correctly identified with ~ 91% accuracy whilst the non-dysplastic (CP-A) and dysplastic (CP-D) cell lines were only correctly classified correctly 63% and 64% of the time. The largest errors occur between CP-A and CP-D cells, with ~ 36% inaccuracy. The OE19 cells could be inaccurately categorised as any of the other three cell lines but with low probability—HET-1A (4%), CP-A (6%) and CP-D (2%). Supplementary Table S1 provides the precision, recall and F1-scores for these results.

Dysplastic progression within Barrett's oesophagus

Next we focused on the analysis of cell lines representing high-grade Barrett's dysplasia. Identification of measurable differences in the spectral data could provide a promising route to the diagnosis of Barrett's Oesophagus severity in patients.

The PCA-LDA was carried out on the fingerprint region of the dysplastic cell lines CP-B (pink), CP-C (orange) and CP-D (purple), shown in Fig. 5. The first 5 PC loadings in Fig. 5b were responsible for 80% of the variance, with the cumulative variance and variance explained detailed in Fig. 5c. Negative bands in PC1 correlated with the aromatic amino acid content and Cyt C. These bands become positive between PC2 and PC4. Other notable bands present in PC1 are associated with Amide I, Amide III, lipid and collagen content. The PO_2^- DNA-associated band (1089 cm^{-1}) also displays an increase in PC1. The scatterplot, Fig. 5a., shows that the different cells types are well separated by LDA1 and 2. The confusion matrix for this data is shown in Fig. 5d, with precision, recall, and F1-scores given in Supplementary Table S2. The average accuracy was calculated at 92%, with PCA-LDA classification accuracy at 77%.

The PCA-LDA is able to correctly identify each stage of Barrett's dysplasia with accurate detection of CP-B at 80%, CP-C at 93% and CP-D at 91%. With CP-B cells have a 13% chance of being identified as CP-C cells.

This low risk of misclassification identifies a new method of detecting progression within Barrett's dysplasia at a single-cell level, which could allow earlier detection of progression towards cancer within a clinical setting. The results from this work not only highlight the use of single-cell Raman spectroscopy alongside PCA-LDA for the identification of healthy versus cancer cells but also for distinguishing between progressive disease stages within Barrett's Oesophagus dysplasia itself.
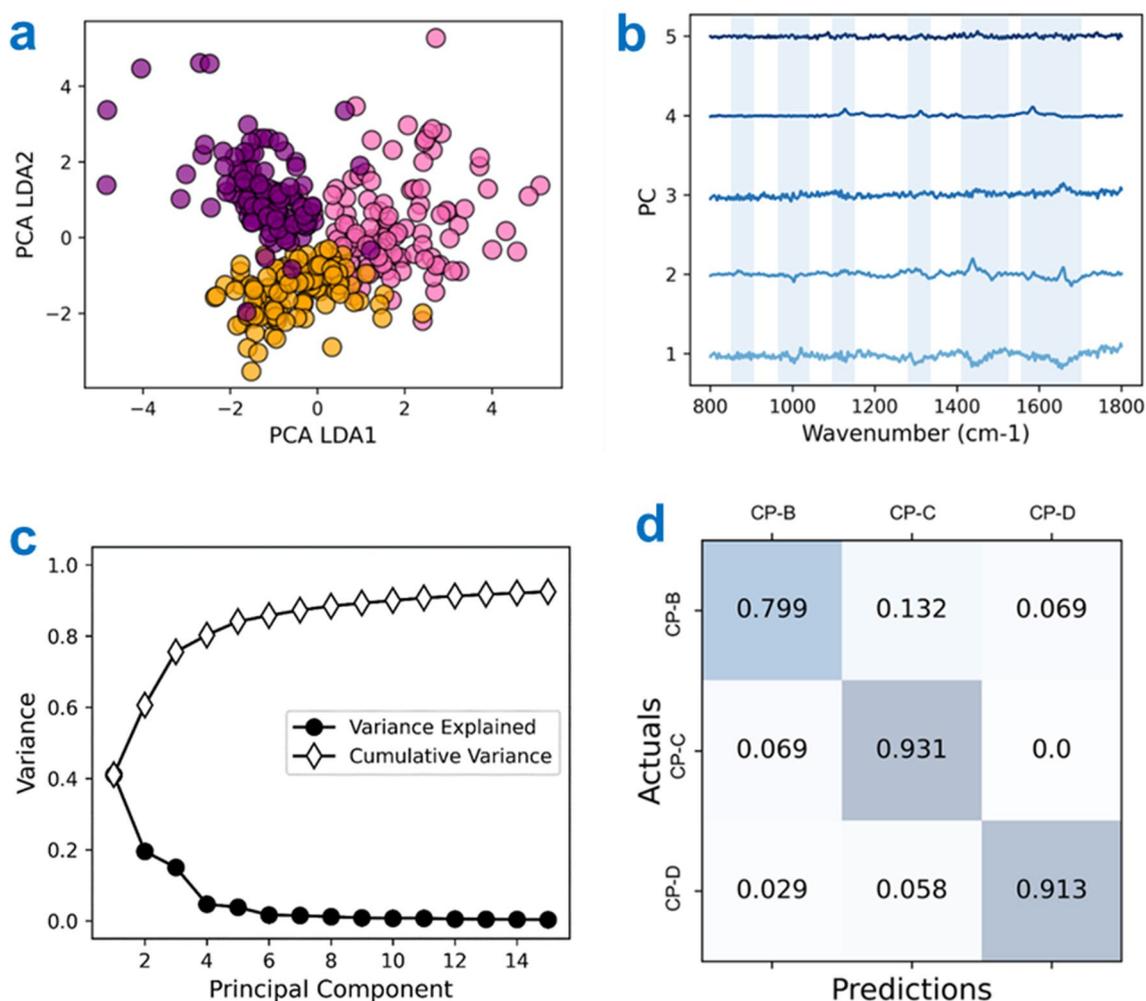

Figure 5. (**a**) 2D plotted PCA LDA1 vs PCA LDA2 scores for CP-B (pink), CP-C (orange) and CP-D (purple). (**b**) Principal component loadings taken from PCA. Scores 1–5 are shown, accounting for > 80% variability. Bands of interest are highlighted in blue. (**c**) Cumulative variance and variance explained for the first 15 PCs. (**d**) Confusion matrix for statistical analysis of the PCA-LDA.

Discussion

Raman spectroscopy provides a "fingerprint" of the molecular covalent bonds present within a biological sample and importantly how they change between different disease states. For example, in this research it is shown how healthy oesophageal cells exhibiting progression of Barrett's Oesophagus toward adenocarcinoma have distinct spectral signatures. In the case of the OE19 cell line, peaks associated with changes in the DNA backbone geometry (Fig. 5b) are indicative of phosphate ion interactions found in neoplastic cells[32,38,49], arising due to the erratic accumulation of DNA and RNA in the nuclei of cancer cells[23].

The increase of the aromatic amino acids associated with advanced disease progression is of note, with previous papers also highlighting the influence of these compounds in various cancers and diseases[35,50]. These compounds have been shown to play a significant role in orchestrating morphological changes within malignant tissues[51]. Moreover, the increase in transcription and replication of cancerous cells requires a concomitantly increased volume of proteins for sustaining the excessive RNA synthesis. This increase in protein expression during cancer progression is expected due to an overall increase in protein overexpression with multiple oncogenic pathway activation and upregulation[25,38,39], as shown in Fig. 2.

Increase in cellular lipid quantity across cell lines (Fig. 2) are significant biomarkers of abnormal membrane composition, characteristic of neoplastic cells. These increases occur due to the modulation of specific enzymes and pathways within the cells, allowing for a higher proliferation rate, increased chemotherapy resistance and attenuation of the immune response[45].

The use of microfluidic devices to trap single cells significantly reduces the volume of reagents required during the treatment and staining of cells. Further, the regular array structure permits the facile and reproducible location of the same cells for multiple measurements over long time courses. In principle, the trapping of cells on a regular array also permits automated cell analysis. This approach offers several potential advantages (i) it could remove the need to culture patient samples, as only a small number of cells are required for testing, and (ii) it is compatible with emerging techniques of sample collection[52,53], which could remove the need for invasive biopsies. The microfluidic platform can be utilised for dissociated cells from patient tissues, or representative cell lines, to allow studies of tumour heterogeneity and/or responses to drug or environmental changes. Our choice of a 60 s acquisition time was a compromise between obtaining a high signal:noise ratio and the time taken to obtain spectra. This could be reduced but would need to be done in conjunction with a study on potential adverse effects on classification accuracy. Alternative approaches, such as Stimulated Raman Spectroscopy, could decrease acquisition times, but the potential compromise associated with the reduction in spectral resolution would need to be considered.

The use of vibrational spectroscopy for quantitative biochemical cellular analysis may provide a non-invasive detection method of disease to allow the classification and grading of dysplasia and malignant neoplasms. The ability to differentiate between these stages has the potential to advance clinical observation of Barrett's Oesophagus progression, potentially allowing earlier intervention and more frequent check-ups for Barrett's Oesophagus patients who are deemed at high risk for cancer development. Within a clinical and research setting, each stage of Barrett's Oesophagus dysplasia is often difficult to distinguish from each other. For continuation of this work, patient samples with a known diagnosis would help in verifying this method as a possible diagnostic tool within a clinical setting. These results highlight not only the potential of combining microfluidic devices with Raman spectroscopy allowing for accurate and sensitive biochemical quantification of disease progression, but also the potential to achieve this using single cells.

Materials and methods

Cell lines

Six cell lines were acquired for these experiments. Human oesophageal cell line HET-1A (ATCC® CRL-2692™) was used for the healthy control. The Barrett's oesophagus in vitro model comprised of one non-dysplastic cell line and three high-grade dysplastic cell lines: CP-A/KR-42421 (ATCC® CRL-4027™), CP-B/ CP-52731 (ATCC® CRL-4028™), CP-C/CP-94251 (ATCC® CRL-4029™) and CP-D/CP-18821 (ATCC® CRL-4030™), respectively. The cell line OE19/JROECL19 (ECACC 96071721) was used as a representative for oesophageal adenocarcinoma. All cell lines were authenticated and confirmed negative for mycoplasma infection.

Cell culture

Healthy (HET-1A) and Barrett's Oesophagus cell lines (CP-A, CP-B, CP-C and CP-D) were cultured in keratinocyte serum-free media (KSFM, Invitrogen), and the oesophageal adenocarcinoma cell line (OE19) was cultured in RPMI cell media containing 2 mM glutamine (Thermo Fisher, UK). Both cell media were supplemented with 10% foetal bovine serum (Merck Life Science Limited). Cells were washed and passaged using Dulbecco's Phosphate Buffered Saline (Merck Life Science Limited) and TrypLE Express Enzyme (Thermo Fisher, UK). Culture was maintained at conditions of 37 °C and 5% CO_2 with a maximum passage number of 20. Cells were passaged immediately before experimentation and re-suspended in their appropriate culture medium before flowing through the respective microfluidic device at a flow rate of 3 μL/min.

Cellular viability on-chip was determined by running both a live and dead stain through the device at 3 μL/ min after each experiment (Viability/Cytotoxicity Assay Kit, Calcein AM, Sytox Red—L32250, Thermo Fisher, UK). A 488 nm optically pumped semiconductor laser aided sequential excitation and fluorescence emission was measured from 490 to 525 nm. Images were captured using a 10× dry objective.

Polydimethylsiloxane (PDMS) microfluidic device fabrication

PDMS (Sylgard, 184, Dowsil) base: curing agent was mixed in a 10:1 ratio before pouring and curing over the designed master wafer. Oxygen plasma bonding (100 W, O2 pressure 0.5 mbar, 30 s, Zepto Plasma Unit, Diener

Electronic, Germany) was used to bind each PDMS device to a 50×20 mm glass coverslip. Devices were sterilised and washed prior to experimentation by pumping through 70% ethanol and Mill-Q consecutively via the use of a PHD ULTRA advanced syringe pump (Harvard Apparatus, USA) at a flow rate of 100 μL/min.

Measurements were taken using an integrated Raman system consisting of an inVia Raman confocal inverted microscope (Renishaw) and a Leica DMi8/SP8 laser scanning confocal microscope. A 532 nm DPSS diode laser, supplying a laser power of 22 mW, at the sample surface, was used with a grating of 1800 lines mm^{-1}. A near-infrared enhanced CCD array detector (1024×256 pixels, Renishaw) was used to collect light. The Raman system was calibrated to the Si band position (520.5 cm^{-1}) prior to each experiment. To ensure no adverse effect due to irradiation, 12 consecutive spectra were obtained from the same "live" cell, these showed no compromise in cellular viability (morphological changes, or spectral differences).

All single-cell spectral data were obtained with the use of a $100 \times$ oil objective (HC PL APO CS2 FWD 0.13 mm NA 1.4. A beam expander was positioned in the optical path between the excitation laser and the objective lens used to focus on the sample. This was used to defocus the laser spot via the Renishaw WiRE software, with 0% defocus (representing the diffraction-limited spot size of the laser), and 100% defocus producing a maximum spot size (dependent on the objective). A 20% defocus, when used with a $100 \times$ objective, produced a spot size of approximately 18 μm at the sample. The entrance slit was set to 20 μm. Spectral settings remained the same across all six cell lines, utilising a step configuration with a 5 s exposure time and 6 accumulations in two different windows (300–1800 cm^{-1} and 1800–3200 cm^{-1}), giving a total exposure time of 60 s per cell. Single Raman spectra was taken from > 100 cells for each cell line. The exact number of single-cell spectra per cell line are as follows: HET-1A (117), CP-A (104), CP-B (109), CP-C (114), CP-D (102), and OE19 (118). Background spectra was acquired at the same Z-position as the cells.

Raman data pre-processing

Individual spectra were compiled into datasets corresponding to each cell line using Matlab's Statistics and Machine Learning Toolbox (MathWorks) with the silicon calibration band for each experiment used to calibrate the wavenumber axis of each spectrum. PDMS background contributions were removed from the data using a modified extended multiplicative signal correction (EMSC) technique[52]. An example of EMSC-corrected CP-A spectra is shown in Supplementary Fig. S4. Dataset files were exported to Python where the outliers with the highest variance were removed to give a final dataset of approximately 100 single-cell spectra per cell type.

All spectra was baseline corrected using the algorithm developed by Koch et al.[53] and truncated to only consider the fingerprint region between 800 and 1800 cm^{-1}. To allow for comparison against other biochemical Raman literature, the spectra were normalised to the Amide I band at 1656 cm^{-1}. This was done for each cell line. Other pre-processing steps included smoothing [polynomial 2, window 17], averaging and calculation of the standard deviation.

PCA-LDA single cell analysis

Principal component analysis and linear discriminant analysis (PCA-LDA) was conducted using the EMSC-corrected datasets shown in Fig. 2, truncated within the fingerprint region (800–1800 cm^{-1}). All confusion matrices presented from this data are normalised using the fourfold cross-validation method.15 principal components were chosen for the PCA as this was representative of > 90% of the overall variance within the data. The variance explained reaches a plateau after 15 PCs, indicating that the remaining components are representative of little to no change. The fourfold cross-validation was conducted using a StratifiedKFold, whereby the datasets for each set of analysis were split into k (4) groups of equal size, with the first group held as a validation set, and the remaining groups (k − 1) used to fit the method.

Data availability

The data presented in this study's findings will be available at https://doi.org/10.5518/1246.

References

1. Hiremath, G., Locke, A., Sivakumar, A., Thomas, G. & Mahadevan-Jansen, A. Clinical translational application of Raman spectroscopy to advance Benchside biochemical characterization to bedside diagnosis of esophageal diseases. *J. Gastroenterol. Hepatol.* **34**, 1911–1921. https://doi.org/10.1111/jgh.14738 (2019).
2. Nowicki-Osuch, K. *et al.* Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. **373**, 760–767. https://doi.org/10.1126/science.abd1449 (2021).
3. Lagergren, J., Smyth, E., Cunningham, D. & Lagergren, P. J. T. L. *Oesophageal Cancer.* **390**, 2383–2396 (2017).
4. Regmi, S., Poudel, C., Adhikari, R. & Luo, K. Q. Applications of microfluidics and organ-on-a-chip in cancer research. *Biosensors* **12**, 1. https://doi.org/10.3390/bios12070459 (2022).
5. Paknahad, A. A., Kerr, L., Wong, D. A., Kolios, M. C. & Tsai, S. S. H. Biomedical nanobubbles and opportunities for microfluidics. *RSC Adv.* **11**, 32750–32774. https://doi.org/10.1039/d1ra04890b (2021).
6. Guo, Q. R. *et al.* Multifunctional microfluidic chip for cancer diagnosis and treatment. *Nanotheranostics* **5**, 73–89. https://doi.org/10.7150/ntno.49614 (2021).
7. Hou, H. W. *et al.* Deformability study of breast cancer cells using microfluidics. *Biomed. Microdev.* **11**, 557–564. https://doi.org/10.1007/s10544-008-9262-8 (2009).
8. Armistead, F. J., Gala De Pablo, J., Gadêlha, H., Peyman, S. A. & Evans, S. D. Cells under stress: An inertial-shear microfluidic determination of cell behavior. *Biophys. J.* **116**, 1127–1135. https://doi.org/10.1016/j.bpj.2019.01.034 (2019).
9. Liu, Y., Li, R., Zhang, L. & Guo, S. Nanomaterial-based immunocapture platforms for the recognition, isolation, and detection of circulating tumor cells. *Front. Bioeng. Biotechnol.* **10**, 850241. https://doi.org/10.3389/fbioe.2022.850241 (2022).

10. Rahmanian, M. *et al.* A micropillar array-based microfluidic chip for label-free separation of circulating tumor cells: The best micropillar geometry?. *J. Adv. Res.* https://doi.org/10.1016/j.jare.2022.08.005 (2022).

11. Ko, J. *et al.* Tumor spheroid-on-a-chip: A standardized microfluidic culture platform for investigating tumor angiogenesis. *Lab. Chip* **19**, 2822–2833. https://doi.org/10.1039/C9LC00140A (2019).

12. Hao, S. *et al.* A spontaneous 3D bone-on-a-chip for bone metastasis study of breast cancer cells. *Small (Weinheim an der Bergstrasse, Germany)* **14**, e1702787. https://doi.org/10.1002/smll.201702787 (2018).

13. Terrell-Hall, T. B., Ammer, A. G., Griffith, J. I. & Lockman, P. R. Permeability across a novel microfluidic blood-tumor barrier model. *Fluids Barriers CNS* **14**, 3. https://doi.org/10.1186/s12987-017-0050-9 (2017).

14. Fang, T. *et al.* Nondestructive identification and accurate isolation of single cells through a chip with Raman optical tweezers. *Anal. Chem.* **91**, 9932–9939. https://doi.org/10.1021/acs.analchem.9b01604 (2019).

15. Ge, S. *et al.* Pump-free microfluidic chip based laryngeal squamous cell carcinoma-related microRNAs detection through the combination of surface-enhanced Raman scattering techniques and catalytic hairpin assembly amplification. *Talanta* **245**, 123478. https://doi.org/10.1016/j.talanta.2022.123478 (2022).

16. Zhang, H., Xiao, L., Li, Q., Qi, X. & Zhou, A. Microfluidic chip for non-invasive analysis of tumor cells interaction with anti-cancer drug doxorubicin by AFM and Raman spectroscopy. *Biomicrofluidics* **12**, 024119. https://doi.org/10.1063/1.5024359 (2018).

17. Wang, Y. *et al.* Microfluidic Raman biochip detection of exosomes: A promising tool for prostate cancer diagnosis. *Lab. Chip* **20**, 4632–4637. https://doi.org/10.1039/D0LC00677G (2020).

18. Canetta, E. Current and FuTURE Advancements of Raman spectroscopy techniques in cancer nanomedicine. *Int. J. Mol. Sci.* **22**. https://doi.org/10.3390/ijms222313141 (2021).

19. Cui, S., Zhang, S. & Yue, S. Raman spectroscopy and imaging for cancer diagnosis. *J. Healthc. Eng.* **2018**, 8619342. https://doi.org/10.1155/2018/8619342 (2018).

20. Santos, I. P. *et al.* Raman spectroscopy for cancer detection and cancer surgery guidance: translation to the clinics. *Analyst* **142**, 3025–3047. https://doi.org/10.1039/c7an00957g (2017).

21. Stomp-Agenant, M. *et al.* In vivo Raman spectroscopy for bladder cancer detection using a superficial Raman probe compared to a nonsuperficial Raman probe. *J. Biophoton.* **15**, e202100354. https://doi.org/10.1002/jbio.202100354 (2022).

22. McGregor, H. C. *et al.* Development and in vivo test of a miniature Raman probe for early cancer detection in the peripheral lung. *J. Biophoton.* **11**, e201800055. https://doi.org/10.1002/jbio.201800055 (2018).

23. Stone, N., Kendall, C., Smith, J., Crow, P. & Barr, H. Raman spectroscopy for identification of epithelial cancers. *Faraday Discuss* **126**, 141–157; discussion 169–183. https://doi.org/10.1039/b304992b (2004).

24. Deng, B., Wang, H., Tan, Z. & Quan, Y. Microfluidic cell trapping for single-cell analysis. *Micromachines* **10**,. https://doi.org/10.3390/mi10060409 (2019).

25. Maitra, I. *et al.* Establishing spectrochemical changes in the natural history of oesophageal adenocarcinoma from tissue Raman mapping analysis. *Anal. Bioanal. Chem.* **412**, 4077–4087. https://doi.org/10.1007/s00216-020-02637-1 (2020).

26. Maitra, I. *et al.* Raman spectral discrimination in human liquid biopsies of oesophageal transformation to adenocarcinoma. *J. Biophoton.* **13**, e201960132. https://doi.org/10.1002/jbio.201960132 (2020).

27. Hao, J. *et al.* The efficacy of Raman spectroscopy in the diagnosis of esophageal cancer: A systematic review and meta-analysis. *Transl. Cancer Res.* **9**, 4750–4761. https://doi.org/10.21037/tcr-20-854 (2020).

28. Carlo, D. D., Wu, L. Y. & Lee, L. P. Dynamic single cell culture array. *Lab. Chip* **6**, 1445–1449. https://doi.org/10.1039/B605937F (2006).

29. Nilsson, J., Evander, M., Hammarström, B. & Laurell, T. Review of cell and particle trapping in microfluidic systems. *Anal. Chim. Acta* **649**, 141–157. https://doi.org/10.1016/j.aca.2009.07.017 (2009).

30. Kuhar, N., Sil, S., Verma, T. & Umapathy, S. Challenges in application of Raman spectroscopy to biology and materials. *RSC Adv.* **8**, 25888–25908. https://doi.org/10.1039/C8RA04491K (2018).

31. Miura, T., Takeuchi, H. & Harada, I. Characterization of individual tryptophan side chains in proteins using Raman spectroscopy and hydrogen-deuterium exchange kinetics. *Biochemistry* **27**, 88–94. https://doi.org/10.1021/bi00401a015 (1988).

32. Shipp, D., Sinjab, F. & Notingher, I. Raman spectroscopy: Techniques and applications in the life sciences. *Adv. Opt. Photon.* **9**, 315. https://doi.org/10.1364/AOP.9.000315 (2017).

33. Stone, N., Kendall, C., Shepherd, N., Crow, P. & Barr, H. *Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers.* **33**, 564–573. https://doi.org/10.1002/jrs.882 (2002).

34. Talari, A. C. S., Movasaghi, Z., Rehman, S. & Rehman, I. u. Raman Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* **50**, 46–111. https://doi.org/10.1080/05704928.2014.923902 (2015).

35. Contorno, S., Darienzo, R. E. & Tannenbaum, R. Evaluation of aromatic amino acids as potential biomarkers in breast cancer by Raman spectroscopy analysis. *Sci. Rep.* **11**, 1698. https://doi.org/10.1038/s41598-021-81296-3 (2021).

36. Puccetti, P. *et al.* Accumulation of an endogenous tryptophan-derived metabolite in colorectal and breast cancers. *PloS one* **10**, e0122046. https://doi.org/10.1371/journal.pone.0122046 (2015).

37. Paul, M. K. & Mukhopadhyay, A. K. Tyrosine kinase—Role and significance in Cancer. *Int. J. Med. Sci.* **1**, 101–115. https://doi.org/10.7150/ijms.1.101 (2004).

38. Chan, J. W. *et al.* Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophys. J.* **90**, 648–656. https://doi.org/10.1529/biophysj.105.066761 (2006).

39. Bergholt, M. S. *et al.* In vivo diagnosis of esophageal cancer using image-guided Raman endoscopy and biomolecular modeling. *Technol. Cancer Res. Treatm.* **10**, 103–112. https://doi.org/10.7785/tcrt.2012.500185 (2011).

40. Xu, Z. *et al.* Application of a microfluidic chip-based 3D co-culture to test drug sensitivity for individualized treatment of lung cancer. *Biomaterials* **34**, 4109–4117. https://doi.org/10.1016/j.biomaterials.2013.02.045 (2013).

41. Cheng, W. T., Liu, M. T., Liu, H. N. & Lin, S. Y. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. *Microsc. Res. Tech.* **68**, 75–79. https://doi.org/10.1002/jemt.20229 (2005).

42. Movasaghi, Z., Rehman, S. & Rehman, I. U. J. A. S. R. *Raman spectroscopy of biological tissues.* **42**, 493–541 (2007).

43. Kaminaka, S., Ito, T., Yamazaki, H., Kohda, E. & Hamaguchi, H.-O. Near-infrared multichannel Raman spectroscopy toward real-time in vivo cancer diagnosis. *J. Raman Spectrosc.* **33**, 498–502. https://doi.org/10.1002/jrs.903 (2002).

44. Malini, R. *et al.* Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: A Raman spectroscopy study. **81**, 179–193 (2006).

45. Szlasa, W., Zendran, I., Zalesińska, A., Tarek, M. & Kulbacka, J. Lipid composition of the cancer cell membrane. *J. Bioenerg. Biomemb.* **52**, 321–342. https://doi.org/10.1007/s10863-020-09846-4 (2020).

46. Overholtzer, M. H., Yakowec, P. S. & Cameron, V. The effect of amino acid substitutions in the conserved aromatic region of subunit II of cytochrome c oxidase in Saccharomyces cerevisiae. *J. Biol. Chem.* **271**, 7719–7724. https://doi.org/10.1074/jbc.271.13.7719 (1996).

47. Rygula, A. *et al. Raman spectroscopy of proteins: a review.* **44**, 1061–1076. https://doi.org/10.1002/jrs.4335 (2013).

48. Eley, C. G. *et al.* Structural role of the tyrosine residues of cytochrome c. *Biochem. J.* **205**, 153–165. https://doi.org/10.1042/bj2050153 (1982).

49. Kendall, C. *et al.* Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *J. Pathol.* **200**, 602–609. https://doi.org/10.1002/path.1376 (2003).

50. Abramczyk, H., Surmacki, J., Brozek-Pluska, B. & Kopec, M. Revision of commonly accepted warburg mechanism of cancer development: redox-sensitive mitochondrial cytochromes in breast and brain cancers by Raman imaging. *Cancers (Basel)* **13**, 2599. https://doi.org/10.3390/cancers13112599 (2021).
51. McDermott, E. W., Barron, E. T., Smyth, P. P. & O'Higgins, N. J. Premorphological metabolic changes in human breast carcinogenesis. *Br. J. Surg.* **77**, 1179–1182. https://doi.org/10.1002/bjs.1800771029 (1990).
52. Ibrahim, O. *et al.* Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections. *Anal. Methods* **9**, 4709–4717. https://doi.org/10.1039/C6AY03308C (2017).
53. Koch, M., Suhr, C., Roth, B. & Meinhardt-Wollweber, M. *Iterative morphological and mollifier-based baseline correction for Raman spectra.* **48**, 336–342. https://doi.org/10.1002/jrs.5010 (2017).

Acknowledgements

Author contributions

A.F. performed the experimental research and data analysis. S.D.E. wrote the code for Raman spectral manipulation and PCA/LDA. A.F, J.E.L, C.D.W and S.D.E. developed the experimental plan, and contributed to the analysis and writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-52079-3.

Correspondence and requests for materials should be addressed to S.D.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.