

# Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

## ***Ex vivo* Raman spectroscopy mapping of lung tissue: label-free molecular characterization of nontumorous and cancerous tissues**

Manon Bourbousson  
Irshad Soomro  
David Baldwin  
Ioan Notingher

**SPIE.**

Manon Bourbousson, Irshad Soomro, David Baldwin, Ioan Notingher, “*Ex vivo* Raman spectroscopy mapping of lung tissue: label-free molecular characterization of nontumorous and cancerous tissues,” *J. Med. Imag.* **6**(3), 036001 (2019), doi: 10.1117/1.JMI.6.3.036001.

# Ex vivo Raman spectroscopy mapping of lung tissue: label-free molecular characterization of nontumorous and cancerous tissues

Manon Bourbousson,<sup>a</sup> Irshad Soomro,<sup>b</sup> David Baldwin,<sup>c</sup> and Ioan Notingher<sup>a,\*</sup>

<sup>a</sup>University of Nottingham, School of Physics and Astronomy, Nottingham, United Kingdom

<sup>b</sup>Nottingham University Hospitals NHS Trust, Histopathology Department, Nottingham, United Kingdom

<sup>c</sup>Nottingham University Hospitals NHS Trust, Department of Respiratory Medicine, Nottingham, United Kingdom

**Abstract.** Raman spectroscopy mapping was used to study *ex vivo* fresh lung tissues and compare to histology sections. The Raman mapping measurements revealed differences in the molecular composition of normal lung tissue, adenocarcinoma, and squamous cell carcinoma (SCC). Molecular heterogeneity of the tissue samples was well captured by the *k*-means clustering analysis of the Raman datasets, as confirmed by the correlation with the adjacent haematoxylin and eosin (H&E) stained tissue sections. The results indicate that the fluorescence background varies considerably even in samples that appear structurally uniform in the H&E images, both for normal and tumor tissue. The results show that characteristic Raman bands can be used to discriminate between tumorous and nontumorous lung tissues and between adenocarcinoma and SCC tissues. These results indicate the potential to develop Raman classifications models for lung tissues based on the Raman spectral differences at the microscopic level, which can be used for tissue diagnosis or treatment stratification.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JMI.6.3.036001]

Keywords: Raman spectroscopy; mapping; lung cancer; diagnosis.

Paper 19090R received Apr. 2, 2019; accepted for publication Jul. 16, 2019; published online Aug. 9, 2019.

## 1 Introduction

Lung cancer represents a major health concern with a ~13% to 17% five-year survival rate.<sup>1</sup> In 2018, lung cancer was, for both sexes combined, the leading cause of cancer deaths worldwide.<sup>2</sup> Where lung cancer is suspected, initial clinical screening tests are performed, including chest x-ray and computed tomography, which is used to both confirm the suspicion and plan further investigations to guide treatment. Diagnosis and tumor staging are performed using a variety of clinical techniques, most commonly a form of biopsy or needle biopsy, using the least invasive and safest approach. Occasionally diagnosis is made during curative surgery, usually where the probability of cancer is very high.<sup>3</sup> The gold standard in diagnosis is the examination of a biopsy of sufficient quality to allow tumor subtyping and molecular analysis. One of the main limitations of histopathology is that it requires a good quality cytology or biopsy sample, followed by lengthy tissue preparation (fixation, sectioning, and staining).

Compared to morphology-based imaging, spectroscopic techniques may provide more an objective diagnosis based on molecular analysis of the tissue. A number of spectroscopy techniques have been investigated for *ex vivo* diagnosis of lung cancer, including nuclear magnetic resonance,<sup>4</sup> (auto)fluorescence spectroscopy,<sup>5–7</sup> infrared spectroscopy,<sup>8,9</sup> and Raman spectroscopy.<sup>10,11</sup> Each technique mentioned has its own advantages and disadvantages in terms of spatial resolution, molecular specificity, signal strength, data acquisition speed, and cost.

Raman spectroscopy has the advantage of high molecular specificity<sup>12</sup> and has been demonstrated to achieve high diagnosis

accuracy for a range of tissue types which include breast,<sup>13,14</sup> skin,<sup>15,16</sup> oral tissues,<sup>17,18</sup> larynx,<sup>19</sup> and prostate.<sup>20,21</sup> Furthermore, Raman spectroscopy is well-suited for *in vivo* or *ex vivo* analysis of fresh tissue as it relies on a back-light scattering configuration, and tissue remains unmodified after analysis, allowing histopathological examination. Huang et al.<sup>10</sup> described the first study using near-infrared (NIR) Raman spectroscopy to examine lung cancer tissue. The study demonstrated the presence of many specific differences in the Raman bands collected from healthy and malignant lung tissues with a specificity and sensitivity of 92% and 94%, respectively. More recently, endoscopic Raman probes utilizing also the high-wavenumber region have been reported, enabling label-free real-time detection of early lung cancers.<sup>22,23</sup>

To reduce the laser-induced fluorescence background, 1064-nm excitation wavelength has also been reported.<sup>11,24</sup> Using the same laser wavelength, Min et al.<sup>11</sup> reported spectral differences in the amide I band at 1659 cm<sup>-1</sup> between normal and cancerous fresh lung tissues. Kaminaka et al.<sup>25</sup> reported that Raman bands assigned to collagen at 1666 and 1448 cm<sup>-1</sup> could be used to discriminate between normal and malignant human lung tissues (tissue fixed in formaldehyde). Although the 1064-nm laser significantly lowered the fluorescence background, it increased the acquisition times for the Raman spectra to 5 min. For *in vivo* diagnosis, investigations based on Raman fibre probes have also been reported. These studies developed flexible and thin fiber optic probes that included both laser delivery and collection of Raman light.<sup>26,27</sup> In addition, Raman spectroscopy has been used to investigate stromal adaptations in premetastatic lungs primed by breast cancer in mouse models.<sup>28</sup>

Although these studies demonstrated the feasibility of Raman spectroscopy for diagnosis of lung cancers, the techniques used to record Raman spectra relied on acquiring single or a low number of Raman spectra from selected regions of the

\*Address all correspondence to Ioan Notingher, E-mail: [ioan.notingher@nottingham.ac.uk](mailto:ioan.notingher@nottingham.ac.uk)

tissue. Because of uncertainties related to the small sampling volume of these Raman probes, single-point measurements can lead to sampling errors caused by structural and molecular heterogeneity of the lung tissue. Probes using high-numerical aperture optics for maximize the collection efficiency of the Raman photons provide high molecular contrast of tissue structures but require accurate co-registration with histopathology. On the other hand, probes with larger sampling volumes (millimeter scale) elevate these issues, but limit the spectral contrast between tissue structures. Raman mapping of lung tissue has been reported, whereby hyperspectral images of tissue specimens, acquired at resolutions as high as 10  $\mu\text{m}$ , were compared to histology images to understand the spectral variability at a microscopic level.<sup>29,30</sup> These studies were performed on frozen-sectioned healthy lung tissues and congenital pulmonary airway malformation but did not include lung cancer tissues.

Here we performed Raman spectral mapping measurements of fresh nontumorous and cancer lung tissue and compare the Raman images to histopathology images. The aim of the study is to provide a better understanding of the effect of tissue heterogeneity on the Raman signals and investigate whether the spectral difference between healthy and cancerous tissue can be used for diagnosis.

## 2 Materials and Methods

### 2.1 Collection of Tissue Samples

Lung tissue samples were collected and analyzed at Nottingham University Hospitals NHS Trust. The samples were collected by the Nottingham Health Science Biobank (NHSB) after obtaining patient consent (ethical approval NHSB REC reference 15/NW/0685). All tissue specimens were received fresh in the Histopathology Department within 20 min of surgery. Surgery included left, right, lower, higher, or partial lobectomy and thoracoscopy. A total of 18 lung tissue samples were investigated.

### 2.2 Tissue Preparation and Histopathology Assessment

Within 20 min following excision, samples of fresh lung tissues were cut in two adjacent blocks: a “Raman block” (used for Raman analysis) and a “reference block” (used for reference analysis by paraffin embedded histology). The Raman blocks were stored frozen ( $-20^{\circ}\text{C}$ ) to minimize alteration. Prussian blue ink was used to paint the outer edge of the reference block to maintain tissue orientation relative to the Raman block when performing the sectioning histology. The reference block was then placed in formalin for fixation, then cut into consecutive sections using a microtome, followed by haematoxylin and eosin (H&E) staining. The H&E sections were assessed by an experienced histopathology consultant (I Soomro). The blue ink was a visual marker to confirm that the section is an adjacent section to the surface of the Raman block, enabling us to correlate the Raman spectral maps to specific structures observed in the H&E section. For Raman spectroscopy experiments, the Raman blocks were defrosted and thoroughly rinsed in phosphate-buffered saline (PBS) at room temperature in order to remove superficial blood (avoids thermal damage caused by absorption of laser light). The tissue was oriented with the cut surface facing the laser beam on the quartz slide. To avoid tissue

drying during the Raman measurements, the tissue was maintained moist using PBS.

### 2.3 Raman Spectroscopy

The Raman spectroscopy measurements were carried out using a Raman microspectrometer based on an inverted optical microscope (Nikon Eclipse Ti-U), 785-nm laser (Xtra, Toptica), and microscope stage (Proscan III, Prior). The microscope was connected with an optical fibre (diameter 100  $\mu\text{m}$ ) to a spectrometer with a 600-lines/mm grating (Oriel 77 200, Newport) and deep depletion back-illuminated CCD (Andor iDus DU-401-A-RR-DD). The 785-nm laser was focused on the sample through a 60 $\times$  1.2 NA oil-immersion objective (RiverD International), providing 160-mW laser power at the sample. We estimate that the sampling volume was  $\sim 5$   $\mu\text{m}$  laterally and  $\sim 50$   $\mu\text{m}$  axially. Raman maps were recorded in mapping mode with 20  $\mu\text{m}$  step size, with an acquisition time of 1 s per spectrum. The instrument was calibrated using the Raman bands of 1,4 bis (2-methylstyryl) benzene and 5.4-acetamidophenol. For three lung samples (one with tumor cells and two normal, with emphysema), the histology assessment carried out after the Raman measurements indicated thermal damage caused by laser exposure. All Raman spectra were dominated by a high-background signal caused likely by the formation of carbon material. Therefore, these tissues are not presented in this study. Although all tissue samples were thoroughly rinsed in PBS prior to the Raman measurements, it is likely that the damage was induced by traces of blood or other contaminants with strong absorption at 785 nm.

### 2.4 Data Analysis

After removal of cosmic ray peaks Raman spectra, *k*-means clustering was applied on the dataset, adjusting the number of clusters until the structures in the *k*-means map resembled the structures observed in the H&E images. To highlight the differences in the Raman spectra, further preprocessing was applied: a baseline subtraction (fifth-order polynomial fitting<sup>31</sup>) and zero mean and unity vector normalization was applied to all Raman spectra.<sup>4,32</sup> This was then followed by principal component analysis (PCA).

Given that the tissue shape and size changes during the histology processing (size changes ranged from 15% to 30%), large errors in co-registration of the Raman images and histology were found. Therefore, instead, the following procedure was used. First, the location of the area analyzed by Raman spectroscopy was recorded relative to the whole area of the tissue. By comparing the overall shape of the tissue with the H&E slide, the location of the scanned area on the H&E image was roughly identified. Then, *k*-means analysis was carried out on the Raman dataset by increasing step wise the number of clusters until the morphological features in the *k*-means maps could be identified and correlated with tissue structures observed in the H&E section.

## 3 Results and Discussion

Figure 1 presents a schematic of the measurement protocols and highlights the numbers and types of tissue samples included in the study. The analysis of the Raman spectra of lung tissue revealed that for four out of the 15 tissue samples investigated, the Raman bands were completely covered by a strong laser-induced fluorescence background, at all locations in the Raman

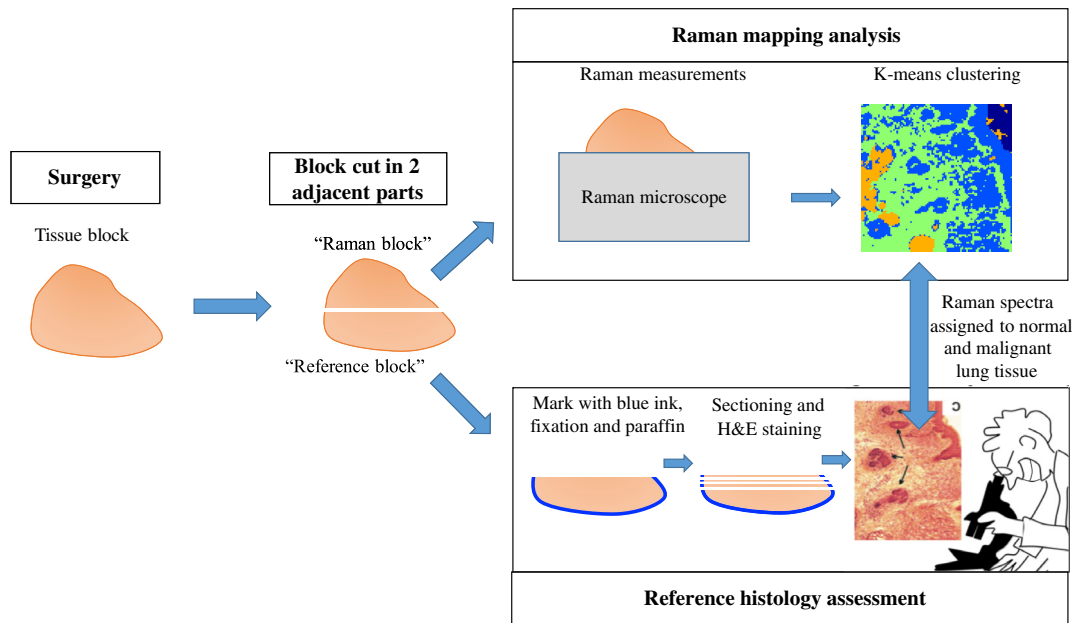


Fig. 1 Schematic description of the tissue preparation steps and study protocol.

maps. Tissue samples with this behavior included both non-malignant and malignant lung tissues. Two examples, one normal and one adenocarcinoma, are presented in Fig. 2. In the case of the adenocarcinoma sample presented in Fig. 2(a), the area of the tissue analyzed consisted of a nest of tumor cells with hyperchromasia. The maximum intensity of the fluorescence emission in the measured spectra ranged from 20,000 counts to higher than 65,000 counts, which is the maximum pixel well of the spectrometer CCD. For the nonmalignant tissue in Fig. 2(b), the H&E indicated that the analyzed region was made of interstitial fibrosis and emphysema. We recorded the maximum fluorescence intensity varying from 4000 counts to 14,000 counts.

For this group of samples, the spectral mapping measurements indicated a high variance in the intensity of fluorescence emission, even for regions of the sample considered structurally homogenous after histopathology evaluation. No particular structure in the tissue could be reliably identified as the source of the laser-induced fluorescence emission.

For the remaining 11 samples, well-defined Raman bands were detected in the measured spectra, despite the presence of a relatively strong fluorescence background, in agreement with previous reports.<sup>10,27</sup> This set of samples included non-malignant lung tissue, squamous cell carcinoma (SCC), and adenocarcinoma.

In total, 17,700 Raman spectra were collected from non-malignant lung tissue samples from three distinct patients. Figure 3 presents typical examples of Raman spectral maps and centroid Raman spectra after *k*-means analysis of the spectral maps.

The spectra in Fig. 3 contain Raman bands that can be assigned to all major groups of biomolecules found in lung tissue.<sup>10,25,27,29</sup> Raman bands corresponding to molecular vibrations in proteins include the protein backbone C–C vibrations (935  $\text{cm}^{-1}$ ), amide I (1655  $\text{cm}^{-1}$ ), amide III (1266 to 1360  $\text{cm}^{-1}$ ). Molecular vibrations of phospholipids lead to intense Raman bands at 1078, 1302, and 1445  $\text{cm}^{-1}$  and some amino acids have characteristic Raman bands, such as phenylalanine (1003  $\text{cm}^{-1}$ ), proline and valine (935  $\text{cm}^{-1}$ ), and tryptophan (1266, 1618  $\text{cm}^{-1}$ ). Additionally, Raman spectra of nucleic acids are typically

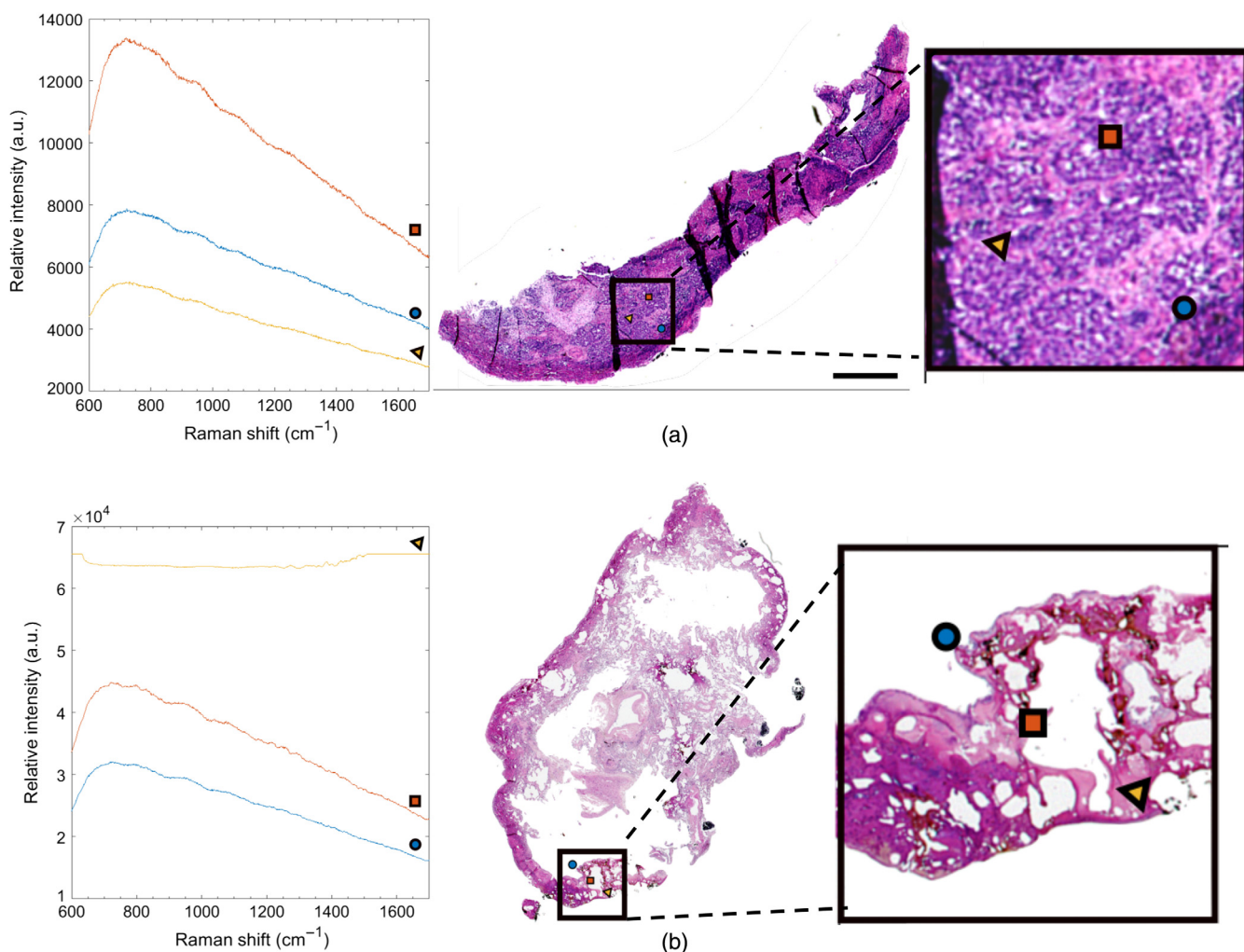
dominated by bands at 788, 1098, 1322, and 1335  $\text{cm}^{-1}$ . Contributions from blood can be identified by the Raman bands at 1370, 1577, and 1620  $\text{cm}^{-1}$ , which were also identified by Krafft et al.<sup>29</sup> and assigned to haemoglobin. Compared to previous studies, we found some differences. For example, we do not observe the Raman band at 1123  $\text{cm}^{-1}$  from proteins mentioned by Huang et al.<sup>10</sup> but we detected a band at 1128  $\text{cm}^{-1}$  instead. Also, in the case of sample (a), additional bands were observed at 1546 and 1605  $\text{cm}^{-1}$  attributed to tryptophan and phenylalanine, respectively.<sup>10</sup>

The pseudocolor images obtained by *k*-means clustering analysis of the Raman measurements confirm the molecular heterogeneity of the samples. For sample (a), which present fibrosis, macrophage, inflamed cells, and fibrosis, the inspection of the centroid Raman spectra indicated significant differences related to both Raman bands and fluorescence emission. Variations in protein content were indicated by differences in band intensities at 855, 935, 1303, and 1252  $\text{cm}^{-1}$  and collagen bands at 935 and 1252  $\text{cm}^{-1}$ .<sup>12</sup> Sample B, characterized by a denser collagen network compared to sample (a), presents emphysema, chronic inflammation, and interstitial fibrosis. The three major clusters generated by the *k*-means accounted for 85% of the area of the Raman map. Although the intensity of the fluorescence emission varied by 50% within the clusters, the main differences in the Raman spectra were observed mainly at 1078, 1128, and 1208  $\text{cm}^{-1}$ , which can be associated to phospholipids, proteins, and specific amino acids, such as tryptophan and phenylalanine.<sup>10,12</sup>

Next, we analyzed five tissue samples from four patients with SCC (total 16,300 Raman spectra). Two typical examples of Raman maps analyzed by *k*-means clustering are presented in Fig. 4. Overall, the spectra of SCC show Raman bands that were also present in the spectra of normal lung tissue, such as bands at 823, 876, 935, 1004, 1123, 1152, 1265, 1302, 1335, 1445, 1552, 1618, and 1655  $\text{cm}^{-1}$ .<sup>10</sup>

The pseudocolor images obtained by *k*-means clustering show that the Raman spectral maps capture the microscopic





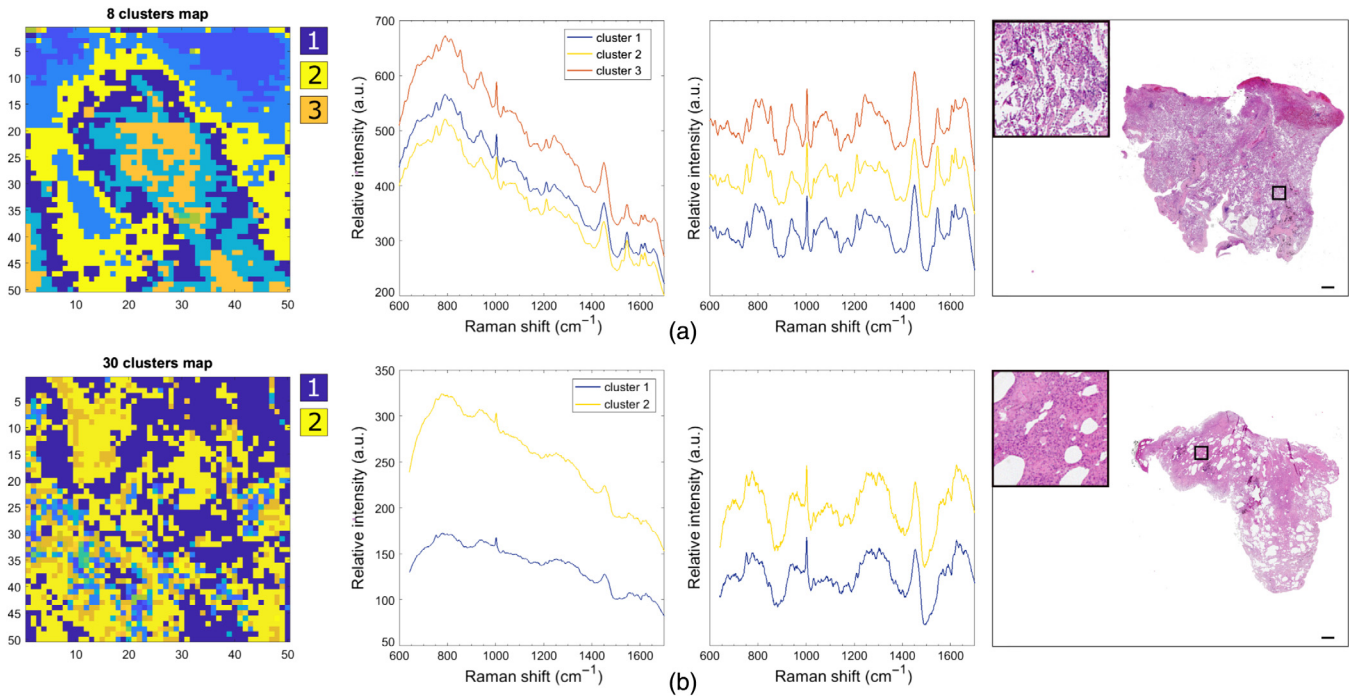
**Fig. 2** Spectral mapping of lung tissues that generated strong laser-induced fluorescence emission that fully swamped the Raman bands. Left panels present typical spectra at three selected positions, indicated by a circle, a square, and a triangle, represented in the corresponding H&E images (right panels). Insets represent the tissue areas analyzed. (a) A sample with adenocarcinoma and (b) nontumorous lung tissue. Scale bar: 1 mm.

molecular and structural heterogeneity of the lung tissue samples well. The H&E images of the two samples indicated that the tissue samples had a variable number of cells within the stroma, which were well discriminated in the Raman spectral maps. Both samples showed poorly differentiated squamous cells. The centroid spectrum corresponding to cluster 2 for sample (a) corresponds to a region of connective tissue, which is distinguishable by stroma fibres and tumor cells. Tissue sample (a) also presents lymphoid follicles in tumor stroma with germinal centers. This region is richer in collagen content compared to the other regions from the same tissue and exhibits a higher level of fluorescence intensity. In contrast, cluster 1 correlates with a region of highly concentrated cells, grouped together. This region is characterized by a spectrum with significantly lower fluorescent intensity (twofold lower). The tissue region in Fig. 4(b) was selected as an area containing a large number of tumor cells and little stroma. The two clusters presented correspond to a high content of tumor cells. The clusters in yellow and green, representing the thin stroma, presented a level of fluorescence background reaching 2760 counts for the peak at  $750\text{ cm}^{-1}$ . The increased fluorescent background is correlated

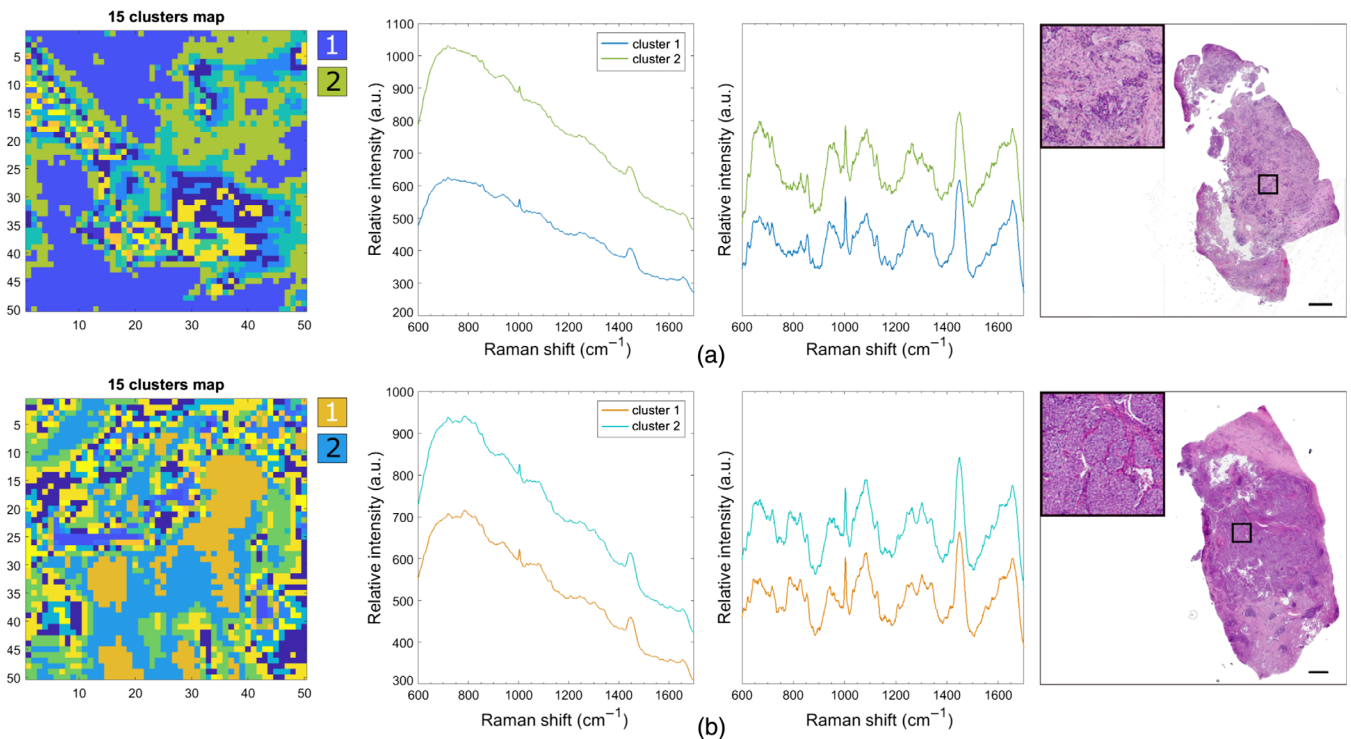
with the lower content of tumor cells. The higher number of cells in this region of sample (b) also led to a higher intensity of the Raman bands assigned to nucleic acids at  $788\text{ cm}^{-1}$  (phosphodiester bond O—P—O stretching vibration) and  $1098\text{ cm}^{-1}$  (phosphodioxy  $\text{PO}_2^-$  group).<sup>12</sup> A higher intensity between the Raman band around  $1090\text{ cm}^{-1}$  was also observed by Huang et al.<sup>10</sup> and assigned to phospholipids. Nevertheless, the direct comparison of the Raman maps and the H&E images allow us to assign this band to nucleic acids and to relate it to SCC regions with higher concentration of tumor cells. This finding agrees with previous studies reporting higher intensity Raman bands of nucleic acids in on other solid tumors (e.g., skin<sup>16</sup> and breast<sup>14</sup>) compared to normal tissue.

Next, we investigated samples containing lepidic adenocarcinoma and adenocarcinoma with necrosis. A total of 18,825 Raman spectra were collected from three adenocarcinoma tissue samples from three patients. Figure 5 presents two typical examples.

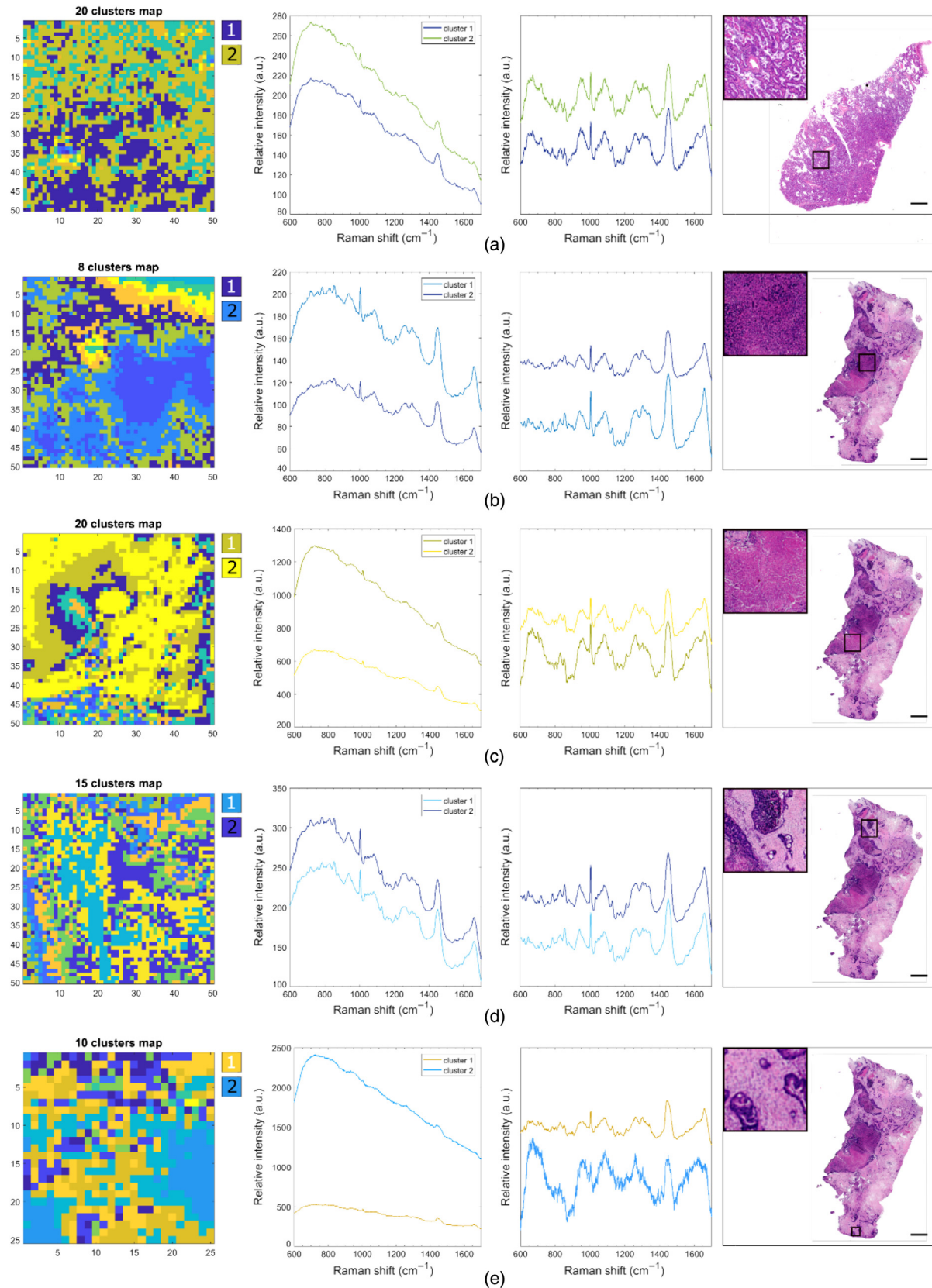
Sample (a) in Fig. 5 presents a lepidic pattern of cancerous gland. The area analyzed by Raman spectroscopy was structurally homogenous. The pseudocolor *k*-means image shows that



**Fig. 3** Spectral mapping of nontumorous lung tissue [two typical samples included (a) and (b)]. Left panels present spectral maps obtained by *k*-means clustering from which centroid spectra with the most intense Raman bands were selected. The spectra in the two middle panels are prior and after preprocessing; the corresponding H&E images are displayed in the right panels (insets represent the tissue areas analyzed). Scale bar: 1 mm.



**Fig. 4** Raman measurement of lung tumor tissues containing SCC. The area highlighted in the microscopic H&E image (right panels) has its *k*-means cluster image represented in the left panels. The average unprocessed Raman spectra alongside the normalized spectra from the mentioned (1 and 2, respectively) clusters are displayed in the two middle panels. Scale bar indicate 1 mm.



**Fig. 5** Raman measurement of lung adenocarcinoma tissues. The tissue in (a) represents lepidic pattern adenocarcinoma; (b)–(e) represent various tissue structures of the same tumor tissue presented in the right panels. The regions consist of: (b) early stage necrosis, (c) late stage necrosis, (d) viable tumour cells, necrosis and stroma, and (e) stroma with viable tumor cells grouped in gland shape. For each, the *k*-means-based image is displayed in the left panel, with corresponding average Raman spectra plotted. Scale bar: 1 mm.



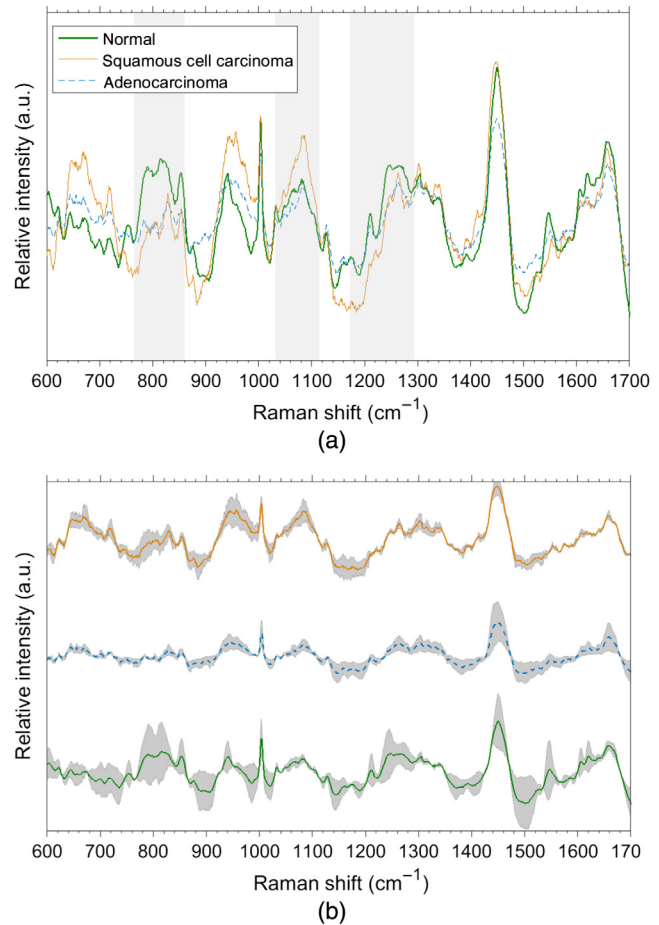
two major clusters accounted for 71% of the analyzed area, for which the centroid spectra contained well-defined Raman bands. These clusters match the pattern observed in the H&E image and represent the tumor cells lining the alveolar walls. Overall, the centroid spectra of tissue (a) shows similar Raman bands and fluorescence background as observed for SCC and normal lung tissue. Nevertheless, one main difference is the  $1209\text{ cm}^{-1}$  band, assigned tentatively to tryptophan and phenylalanine,<sup>32</sup> which is more pronounced compared to SCC. Also there are differences in the shapes of the bands at  $1302$  and  $1335\text{ cm}^{-1}$ . Finally, the Raman band assigned to tryptophan at  $1618\text{ cm}^{-1}$  is more intense for the adenocarcinoma sample (a) compared to the SCC samples.

The second tissue presented in Fig. 5 had a highly heterogeneous structure and several regions were investigated by Raman mapping: early stage necrosis [Fig. 5(b)], late stage necrosis [Fig. 5(c)], a mixed region of viable cells, necrosis and stroma [Fig. 5(d)], and a region of stroma with viable tumor cells grouped in a gland shape [Fig. 5(e)]. For the region of early necrosis [Fig. 5(b)], the centroid spectra for the eight clusters selected the *k*-means had overall similar Raman spectral characteristics, confirming the low level of structural heterogeneity. However, the spectra of this region showed the lowest levels of laser-induced fluorescence compared to all other samples investigated. The Raman dataset acquired from the region of late necrosis was analyzed using 20 clusters. While no significant differences were observed for the Raman bands, strong fluorescence emission (higher than 10,000 counts) dominated 16 centroid spectra, swamping any Raman bands. For the remaining four clusters, accounting for 87% of the Raman map, the fluorescence varied between 700 and 3500 counts and the Raman bands were clearly discernible. The spectra from the late necrosis region presented a band at  $1415\text{ cm}^{-1}$  not assigned in the literature and not present in the other adenocarcinoma tissue structure. Additionally, we observed differences in the shape of some Raman bands in the region  $1550$  to  $1700\text{ cm}^{-1}$ ; this region had more intense and marked bands assigned to porphyrin and tryptophan at  $1518$ ,  $1552$ , and  $1612\text{ cm}^{-1}$ .<sup>12</sup>

For the region containing connective tissue, the *k*-means cluster analysis provides good discrimination between the connective tissue and the regions containing larger numbers of viable cells. Similar to the spectra measured for SCC, the spectra of stroma show a higher fluorescence background, whereas the Raman bands assigned to nucleic acids have higher intensity in the regions with larger number of tumor cells.

Overall, we observed spectral differences between the structures in tissue (b). The bands assigned to protein backbone stretching  $\nu(\text{C}-\text{C})$  and collagen identified at  $942\text{ cm}^{-1}$  for non-malignant and malignant tissues a shifted its maximum at  $958$  and  $956\text{ cm}^{-1}$  for the late necrosis region and mixed region, respectively. The  $1552$  and  $1518\text{ cm}^{-1}$  corresponding to porphyrin vibrations had the highest intensity in the regions of most advanced stage of necrosis, an observation reported in a study that shows that higher contents of porphyrins were observed in necrotic regions of tissues.<sup>4</sup>

Finally, we compared all Raman spectra of nontumorous and cancerous lung tissue to investigate whether the spectral differences between normal and cancerous lung tissues are larger than the inter- and inpatient variance. Figure 6 presents the average Raman spectra for healthy lung tissue, adenocarcinoma and SCC after subtraction of the fluorescence baseline. The largest intensity variations in the Raman spectra



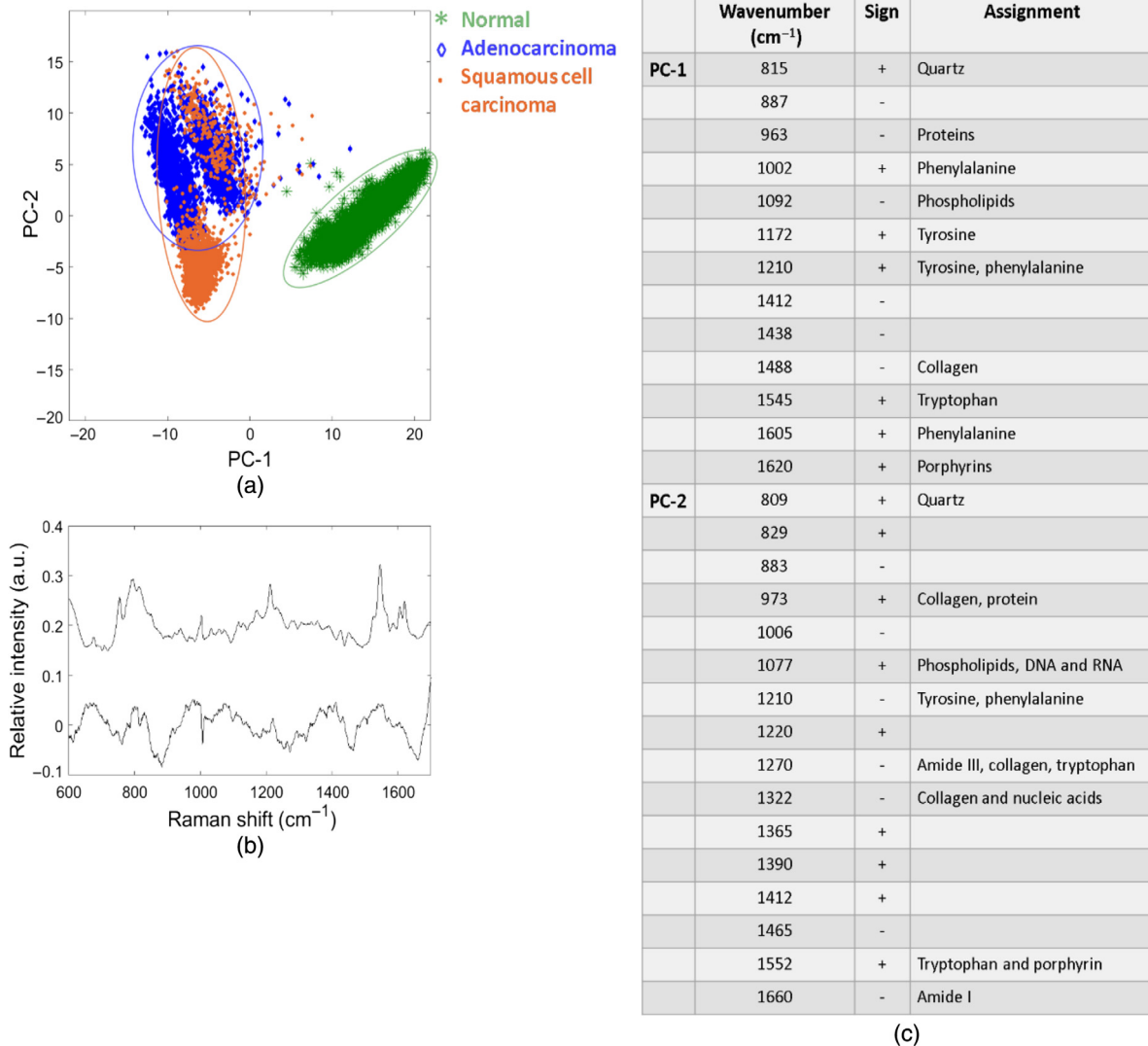
**Fig. 6** (a) Mean Raman spectra of normal lung tissue (green), adenocarcinoma (dotted blue), and squamous cell carcinoma (orange). Each spectrum is the average of around 18,000 spectra from 3 patients (normal), 19,000 spectra from 3 patients (adenocarcinoma), and 16,000 spectra from 5 patients (squamous cell carcinoma). The regions shaded in light gray highlight the main spectral differences between the three types. (b) The shaded regions indicate  $\pm 2$  standard deviation from the mean. The spectra are represented on the same intensity scale and are offset for clarity.

were observed for nontumorous lung tissue. Subtler variations spread over the entire spectral range were observed for SCC and adenocarcinoma. Although most of the spectral variance can be accounted on interpatient variability, variations in fluorescence background can also lead to spectral distortions after the subtraction of the baseline (e.g.,  $800$  to  $900\text{ cm}^{-1}$  region).

The mean spectra of SCC and adenocarcinoma are similar with slight differences in relative intensity of certain bands, such as  $906$  to  $996$ ,  $1031$  to  $1110$ , the  $1445\text{ cm}^{-1}$  assigned to phospholipids and proteins, and peak specific to SCC at  $1394$  and  $1413\text{ cm}^{-1}$ . At the same time, the intensities of bands at  $790$ ,  $814$ , and  $855\text{ cm}^{-1}$  are higher for nontumorous tissues compared to tumor tissues, as also reported previously.<sup>10,25</sup> A similar behavior is observed for the nontumorous samples which have strongest signal for the peaks at  $1248$ ,  $1269\text{ cm}^{-1}$  from amide III, collagen, tryptophan, and phospholipids vibrations and at  $1618\text{ cm}^{-1}$  from tryptophan and porphyrin.

The polynomial curve fitting method used as a baseline subtraction to remove the fluorescence background can lead to spectral artefacts in the cases of high level of fluorescence. To minimize the artefacts, a threshold criterion was implemented





**Fig. 7** (a) PCA of the Raman spectra of nonmalignant lung tissue, adenocarcinoma, and squamous cell carcinoma tissue samples, (b) PC-1 and PC-2 loading, and (c) assignments of the positive (+) and negative (-) peaks appearing in the loading of PC-1 and PC-2.

based on the calculation of the signal-to-noise ratio (SNR) from the peak at  $1450\text{ cm}^{-1}$ . The SNR for a specific Raman band ( $y$ ) is defined as  $\text{SNR} = \frac{\bar{s}}{\sigma_y}$ , where  $\bar{s}$  represent the peak height and  $\sigma_y$  is the standard deviation of the peak height. By discarding the lowest value of SNR ( $\text{SNR} < 2$ ), the spectra with small Raman bands contribution overshadowed in an intense fluorescence background were removed. We applied this method to the database and we obtained a reduced spectral dataset including 2497 spectra from nonmalignant tissue, 2518 spectra from adenocarcinoma, and 2879 spectra from SCC. This dataset was then further analyzed by PCA. Figure 7 presents the scores and the loading of the first two principal components PC-1 and PC-2 which accounted for 24% and 7% of the variance. The results in Fig. 7 show that discrimination between healthy tissue, adenocarcinoma, and SCC can be achieved. The PC-1 versus PC-2 plot in Fig. 7(a) shows two distinct clouds of data points that separate nontumorous and tumorous tissues. Figure 7(b) shows that the loadings of the principal components responsible for the spectral separation between the three types of lung tissue contain complex patterns of positive and negative bands, suggesting that the separation is based on a complex molecular pattern. The nontumorous tissues are

characterized by positive values of PC-1, including Raman bands assigned to phenylalanine ( $1002$  and  $1605\text{ cm}^{-1}$ ), amino acids ( $1172$ ,  $1210$ , and  $1545\text{ cm}^{-1}$ ) and haemoglobin ( $1620\text{ cm}^{-1}$ ). The groups of data from SCC and adenocarcinoma are located in the negative part of PC-1 illustrating the contribution from the peaks from proteins ( $963\text{ cm}^{-1}$ ), phospholipids ( $1092\text{ cm}^{-1}$ ), and collagen ( $1488\text{ cm}^{-1}$ ) in their spectra.

## 4 Conclusions

In this study, we analyzed the Raman spectral maps obtained from fresh normal and malignant lung tissue samples. The Raman mapping measurements revealed differences in the molecular composition of normal lung tissue, adenocarcinoma, and SCC. Molecular heterogeneity of the tissue samples was well captured by the  $k$ -means clustering analysis of the Raman mapping datasets, as confirmed by the correlation with the adjacent H&E stained tissue sections. The results indicated that the fluorescence background varied considerably even in samples that appear structurally uniform in the H&E images, both for normal and tumor tissue.

Experimental data showed that there were characteristic Raman peaks to discriminate between malignant and nonmalignant lung tissues and between adenocarcinoma, and SCC tissues. These results indicate the potential to develop Raman classifications models for lung tissues based on the Raman spectral differences at the microscopic level, which can be used for tissue diagnosis or treatment stratification.

### Disclosures

The authors declare no conflicts of interest.

### Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (Grant No. EP/L025620/1). We also gratefully acknowledge the contributions of the Nottingham Health Services Biobank and the Department of Pathology at Nottingham University Hospitals.

### References

1. S. Walters et al., "Is England closing the international gap in cancer survival?" *Br. J. Cancer* **113**(5), 848–860 (2015).
2. F. Bray et al., "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: Cancer J. Clin.* **68**(6), 394–424 (2018).
3. L. G. Collins et al., "Lung cancer: diagnosis and management," *Am. Fam. Physician* **75**(1), 56–63 (2007).
4. C. M. Rocha et al., "Metabolic profiling of human lung cancer tissue by 1H high resolution magic angle spinning (HRMAS) NMR spectroscopy," *J. Proteome Res.* **9**(1) 319–332 (2009).
5. N. Ramanujam, "Fluorescence spectroscopy of neoplastic and non-neoplastic tissues," *Neoplasia* **2**(1-2), 89–117 (2000).
6. J. Hung et al., "Autofluorescence of normal and malignant bronchial tissue," *Lasers Surg. Med.* **11**(2), 99–105 (1991).
7. A. C. Croce and G. Bottiroli, "Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis," *Eur. J. Histochem.* **58**(4), 2461 (2014).
8. K. Yano et al., "Evaluation of glycogen level in human lung carcinoma tissues by an infrared spectroscopic method," *Cancer Lett.* **110**(1-2), 29–34 (1996).
9. P. D. Lewis et al., "Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum," *BMC Cancer* **10**(1), 640 (2010).
10. Z. Huang et al., "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer* **107**(6), 1047–1052 (2003).
11. Y.-K. Min et al., "1064 nm near-infrared multichannel Raman spectroscopy of fresh human lung tissues," *J. Raman Spectrosc.* **36**(1), 73–76 (2005).
12. D. W. Shipp, F. Sinjab, and I. Notingher, "Raman spectroscopy: techniques and applications in the life sciences," *Adv. Opt. Photonics* **9**(2), 315–428 (2017).
13. A. S. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Nat. Acad. Sci. U. S. A.* **102**(35), 12371–12376 (2005).
14. D. W. Shipp et al., "Intra-operative spectroscopic assessment of surgical margins during breast conserving surgery," *Breast Cancer Res.* **20**(1), 69 (2018).
15. T. C. Bakker-Schut et al., "Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy," *J. Invest. Dermatol.* **119**(1), 64–69 (2002).
16. K. Kong et al., "Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy," *Proc. Nat. Acad. Sci. U. S. A.* **110**(38), 15189–15194 (2013).
17. R. Malini et al., "Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study," *Biopolymers* **81**(3), 179–193 (2006).
18. E. M. Barroso et al., "Water concentration analysis by Raman spectroscopy to determine the location of the tumor border in oral cancer surgery," *Cancer Res.* **76**(20) 5945–5953 (2016).
19. N. Stone et al., "Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers," *J. Raman Spectrosc.* **33**(7), 564–573 (2002).
20. P. Crow et al., "Assessment of fiberoptic near-infrared Raman spectroscopy for diagnosis of bladder and prostate cancer," *Urology* **65**(6), 1126–1130 (2005).
21. P. Crow et al., "The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines," *Br. J. Cancer* **92**(12), 2166–2170 (2005).
22. H. C. McGregor et al., "Real-time endoscopic Raman spectroscopy for in vivo early lung cancer detection," *J. Biophotonics* **10**(1), 98–110 (2017).
23. H. C. McGregor et al., "Development and in vivo test of a miniature Raman probe for early cancer detection in the peripheral lung," *J. Biophotonics* **11**(11), e201800055 (2018)
24. C. A. Patil et al., "1064 nm dispersive Raman spectroscopy of tissues with strong near-infrared autofluorescence," *Opt. Lett.* **39**(2), 303–306 (2014).
25. S. Kaminaka et al., "Near-infrared Raman spectroscopy of human lung tissues: possibility of molecular-level cancer diagnosis," *J. Raman Spectrosc.* **32**(2), 139–141 (2001).
26. M. A. Short et al., "Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers," *Opt. Lett.* **33**(7), 711–713 (2008).
27. N. D. Magee et al., "Ex vivo diagnosis of lung cancer using a Raman miniprobe," *J. Phys. Chem. B* **113**(23), 8137–8141 (2009).
28. S. D. Pardi et al., "Label-Free Raman spectroscopy detects stromal adaptations in premetastatic lungs primed by breast cancer," *Cancer Res.* **77**(2), 247–256 (2017).
29. C. Krafft et al., "Raman and FTIR imaging of lung tissue: methodology for control samples," *Vib. Spectrosc.* **46**(2), 141–149 (2008).
30. S. Koljenovic et al., "Raman microspectroscopic mapping studies of human bronchial tissue," *J. Biomed. Opt.* **9**(6), 1187–1198 (2004).
31. C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.* **57**(11), 1363–1367 (2003).
32. N. Stone et al., "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss.* **126**, 141–157 (2004).

Biographies of the authors are not available.