

Raman microscopy in the diagnosis and prognosis of surgically resected nonsmall cell lung cancer

Nicholas David Magee

The Queen's University of Belfast
Centre for Infection and Immunity
Respiratory Medicine Research Group
Microbiology Building
Grosvenor Road
Belfast, BT12 6BN United Kingdom

James Renwick Beattie

Centre for Vision and Vascular Science
Institute of Clinical Sciences
Grosvenor Road
Belfast, BT12 6BA United Kingdom

Chris Carland

Richard Davis

Royal Victoria Hospital
Department of Pathology
274 Grosvenor Road
Belfast, BT12 6BA United Kingdom

Kieran McManus

Royal Victoria Hospital
Department of Cardiothoracic Surgery
274 Grosvenor Road
Belfast, BT12 6BA United Kingdom

Ian Bradbury

Department of Epidemiology and Public Health
73 University Road
Belfast, BT7 1NN United Kingdom

Dean Andrew Fennell

Peter William Hamilton

Queen's University Belfast
Centre for Cancer Research and Cell Biology
97 Lisburn Road
Belfast, BT9 7BL United Kingdom

Madeleine Ennis

The Queen's University of Belfast
Centre for Infection and Immunity
Respiratory Medicine Research Group
Microbiology Building
Grosvenor Road
Belfast, BT12 6BN United Kingdom

John Joseph McGarvey

Queen's University Belfast
School of Chemistry & Chemical Engineering
David Keir Building
Stranmillis Road
Belfast, BT9 5AG United Kingdom

Joseph Stuart Elborn

The Queen's University of Belfast
Centre for Infection and Immunity
Respiratory Medicine Research Group
Microbiology Building
Grosvenor Road
Belfast, BT12 6BN United Kingdom

Abstract. The main curative therapy for patients with nonsmall cell lung cancer is surgery. Despite this, the survival rate is only 50%, therefore it is important to more efficiently diagnose and predict prognosis for lung cancer patients. Raman spectroscopy is useful in the diagnosis of malignant and premalignant lesions. The aim of this study is to investigate the ability of Raman microscopy to diagnose lung cancer from surgically resected tissue sections, and predict the prognosis of these patients. Tumor tissue sections from curative resections are mapped by Raman microscopy and the spectra analyzed using multivariate techniques. Spectra from the tumor samples are also compared with their outcome data to define their prognostic significance. Using principal component analysis and random forest classification, Raman microscopy differentiates malignant from normal lung tissue. Principal component analysis of 34 tumor spectra predicts early postoperative cancer recurrence with a sensitivity of 73% and specificity of 74%. Spectral analysis reveals elevated porphyrin levels in the normal samples and more DNA in the tumor samples. Raman microscopy can be a useful technique for the diagnosis and prognosis of lung cancer patients receiving surgery, and for elucidating the biochemical properties of lung tumors. © 2010 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3323088]

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

Paper 09187R received May 11, 2009; revised manuscript received Sep. 3, 2009; accepted for publication Dec. 31, 2009; published online Mar. 12, 2010.

1 Introduction

Lung cancer is still the most common cancer-related death globally, with an estimated 1.3 million new cases diagnosed annually.¹ 80% of cases are histologically classified as nonsmall cell lung cancer, for which surgery is the main curative treatment. Despite this, 50% of these patients suffer recurrence within 5 years. A number of prognostic markers that predict postoperative recurrence have been suggested, but

Address all correspondence to: Nicholas Magee, Respiratory Medicine Research Cluster, Respiratory Research Office, Belfast City Hospital, Lisburn Road, Belfast, BT9 7AB, Northern Ireland. Tel: 44-2890-263-729; Fax: 44-2890-263-546; E-mail: nmagee@doctors.org.uk

none have been clinically validated. Several optical techniques are being developed for the diagnosis of lung cancer, the most prominent being autofluorescence.^{2,3} Raman spectroscopy is a well-established analytical technique where the structure and binding of molecules can be studied by examination of their light scattering properties.⁴⁻⁶ Recently, the sensitivity of Raman spectrometers has improved as a result of several technical advances, allowing the acquisition of high quality Raman spectra from tissue sections and cells.⁵ Because no staining or tissue preparation is required, and Raman studies can be conducted under nondestructive conditions, the technique can be applied *in vivo*.^{6,7} The method has been applied to a number of biomedical applications, and several solid organ cancers have been successfully analyzed using Raman spectroscopy.⁸⁻¹² Two studies have recently demonstrated that Raman spectroscopy can differentiate malignant from normal lung tissue.^{13,14} Huang et al. used an excitation wavelength of 785 nm to distinguish tumor from normal bronchial tissue, with a sensitivity and specificity of 94 and 92%, using only the ratio of two distinctive features in the Raman spectrum 1445:1655 cm^{-1} . Yamazaki et al. used a higher wavelength of 1064 nm to distinguish tumor from normal parenchymal tissue, and obtained a sensitivity of 91% and a specificity of 97%. However, the conventional Raman spectroscopy approach employed in these studies analyzed contributions from a heterogeneous tissue involving a complicated combination of cell types and stromal tissues, all with different biochemical profiles. This issue has been addressed in a study examining normal bronchial tissue sections using Raman microscopy, in which conventional Raman spectroscopy is combined with optical microscopy to achieve spatial as well as spectral resolution.¹⁵ Several chemical differences were identified between the individual layers, such as epithelium and supporting stromal layers.

Our aim in the present study has been to investigate whether Raman microscopy can: 1. distinguish lung cancer cells from normal bronchial epithelium, 2. explore its potential for prognosis in patients undergoing lung cancer surgery, and 3. investigate biochemical differences between normal and malignant bronchial cells.

2 Materials and Methods

2.1 Sample Preparation

43 patients undergoing lung cancer resection were recruited for this study. A $1 \times 0.5 \times 0.5 \text{ cm}^3$ sample of tumor tissue and normal lung tissue, distant from the tumor, were obtained from each patient. Of the 43 normal lung samples, 28 contained identifiable normal bronchial epithelium. Of the 43 tumor specimens, 34 were suitable for analysis: four were not nonsmall cell cancer, two tumors were too small for residual tissue to be collected, and three tumor tissues suffered burn damage in the analyzing laser beam. Using a cryotome stage, the sections were mounted in Tissue Tek[®], and frozen tissue sections measuring 5 μm in thickness were prepared. The first section was placed on a glass microscope slide and stained with hematoxylin and eosin. The next consecutive section was placed onto a quartz microscope slide and then immersed for 5 min in 99% ethanol to preserve the tissue. This

second unstained section was then analyzed by Raman microscopy using the first section as a “stained map.” These maps were also assessed by a pathologist and given a diagnosis of normal or nonsmall cell lung carcinoma. All the patients gave informed consent, and the study had the approval of the Queen’s University Belfast Research Ethics Committee.

2.2 Confocal Raman Microscopy

Raman spectra were recorded using a Horiba Jobin Yvon LabRam HR800 Raman microscope (Horiba, Limited, Kyoto, Japan). Light from a 785-nm diode laser was focused on the sample with a 100 \times objective (numerical aperture is 0.9) and an excitation power of 20 mW at the sample. A 785-nm notch filter was used to reject elastic scattering from the incident light and a 300-groove mm^{-1} diffraction grating was used to provide a spectral resolution of 6 cm^{-1} over the spectral range from 600 to 1800 cm^{-1} on a Peltier-cooled charge-coupled device (CCD) detector (Andor Technology, Belfast, Ireland, model DU420). Spectra were acquired and processed using Labspec software (Jobin-Yvon, Villeneuve d’Ascq, France).

Using the stained section of normal bronchial tissue for guidance, two normal areas of bronchial epithelium were identified on the unstained section and scanned with the Raman microscope. These areas measured on average $30 \times 30 \mu\text{m}$, included five to ten cells, and were sampled over 16 points using a 4×4 square grid. The integration time for each point was four minutes. Mean spectra were calculated from these areas; therefore for each patient, two spectra were obtained for normal tissue and two for malignant tissue.

2.3 Spectral Preprocessing and Analysis

Residual cosmic spikes were removed from each spectrum and a mean spectrum was calculated using the 16 spectra from the 4×4 grid. The position of the phenylalanine peak was used to check for any x -axis shift in recorded spectra from sample to sample. Raman scattering from a clean quartz slide was measured in a similar fashion to the tissue and was subtracted from each of the mean spectra to reduce the influence of quartz in the spectra. Background noise and fluorescence were removed by multiple linear baseline subtraction; nine points over regions known to have no major Raman activity were manually chosen to fit the background, and this baseline was subtracted from the spectra. The spectra were then normalized by dividing by the total area under the curve.

Mean normal and malignant spectra were compared using a Student’s t -test. The ability of Raman microscopy to correctly classify normal and malignant lung tissue was tested using principal component analysis (PCA)¹⁶ with a leave-one-out cross-validation (LOOCV), and by Random forest classification¹⁷ with training and test sets. The ability of Raman microscopy to predict postoperative recurrence in 34 patients was also analyzed by principal component analysis with leave-one-out cross-validation. A survival analysis for these patients was performed using the scores from the recurrence analysis and pathological parameters (log-rank test). These statistical analyses were performed by Unscrambler (Camo, Oslo, Norway), SPSS (SPSS Incorporated, Chicago, Illinois), and the R software (R Foundation for Statistical Computing, Vienna, Austria).

Table 1 Patient characteristics.

	Patients with early recurrence	Patients without early recurrence	
Mean age (range)	63 (49 to 79)	66 (51 to 78)	
p-stage status			
1	8	11	
2	4	6	
3	3	2	$p=0.45$
Adjuvant chemotherapy	7/15 (47%)	5/19 (26%)	$p=0.29$
Lymphovascular Invasion	8/15 (53%)	7/19 (37%)	$p=0.49$

2.4 Reference Spectra

To assist with spectral identification of the loadings from the PCA, spectra from several reference materials were recorded with the Raman microscope using 785-nm excitation. Horse heart cytochrome C (C-7150) was purchased from Sigma (Saint Louis, Missouri). Human DNA was isolated from the whole blood of a control subject. Genomic DNA was purified using “salting-out” method with a Gentra Puregene purification kit.

3 Results

Patient characteristics for the 34 patients in the prognostic analysis are shown in Table 1. There were 15 (44%) cases of early postoperative recurrence, defined on the basis of radiological or pathological evidence of local or distant recurrence within 12 months of surgical resection. Recurrence only in the mediastinal nodes was noted in three patients. The remaining recurrences were metastatic.

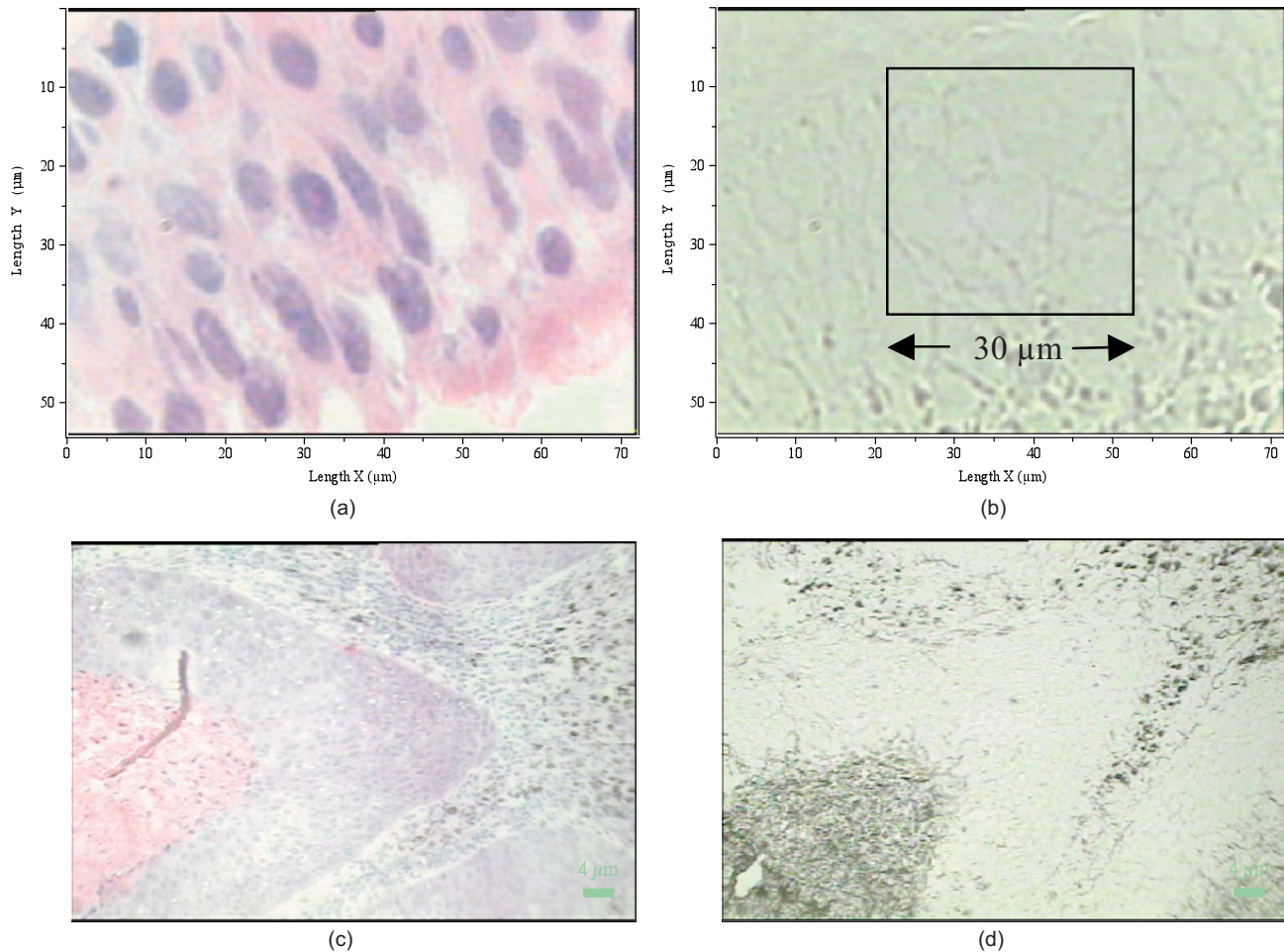


Fig. 1 Light microscopy images of normal samples (63×): (a) is stained with hematoxylin and eosin; the box in (b) illustrates the area of tissue analyzed with Raman microscopy; (c) shows light microscopy images of a malignant sample (10×) stained with hematoxylin and eosin; and (d) shows the consecutive unstained section.

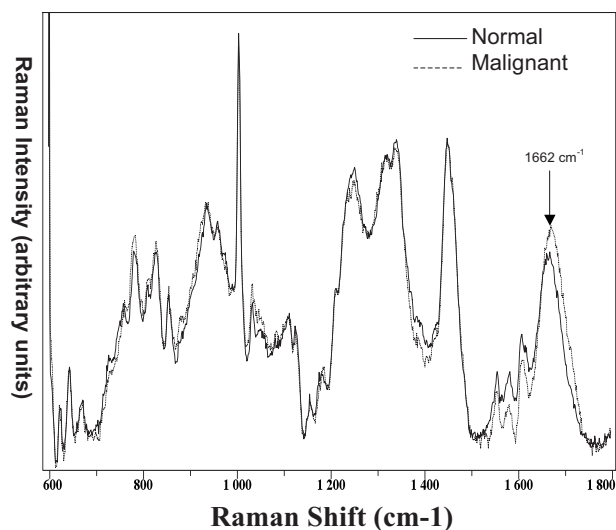


Fig. 2 Mean spectra from normal and malignant samples (after preprocessing).

Light microscopy images of two sequential normal tissue sections are shown in Figs. 1(a) (stained with hematoxylin and eosin) and 1(b) (not stained). The box within Fig. 1(b) illustrates the area analyzed by the Raman microscope. Figures 1(c) and 1(d) show a malignant tissue section at a lower magnification. Again, these are consecutive sections and only Fig. 1(c) has been stained. The histopathological diagnosis was confirmed on the stained section, which was morphologically indistinguishable from the section used for Raman analysis. These figures illustrate that the stained sections can be used as a guidance map for the unstained sections that were analyzed by Raman microscopy. The figure also illustrates how a small group of individual cells can be targeted with the microscope.

There are many subtle differences in intensities between the two main spectra but the main difference identified was in the intensity and width of the amide I band at 1662 cm^{-1} (Fig. 2), which is consistent with reports in previous papers on both lung cancer and other solid organ tumors.^{8–13} However, it should be pointed out that there was no significant difference between the mean intensities at 1662 cm^{-1} when compared using an independent samples t-test ($p=0.894$).

In the principal component analysis of the 124 normal and malignant spectra recorded in the present study, 60% of variation in the spectra was described by the first two principal components. These two components also most efficiently classified normal and tumor spectra, for which a scatter plot is shown in Fig. 3. The dashed, diagonal line separates these score plots into normal and malignant groups. Thus, it was found that Raman microscopy can differentiate malignant from normal lung tissue with a sensitivity of 84% and a specificity of 61% (positive predictive value 72%, negative predictive value 76%).

Principal component loadings are a graphical representation of the mathematical function that PCA uses to explain the separation between normal and tumor spectra. Their interpre-

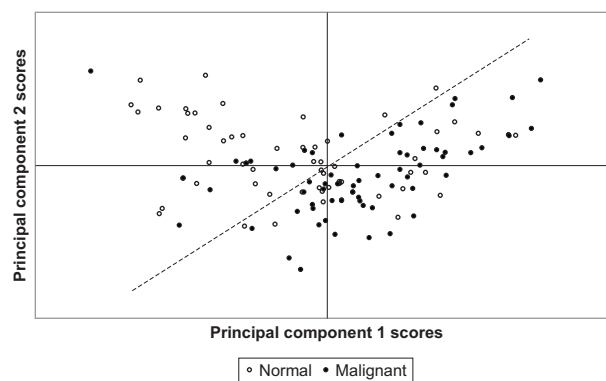


Fig. 3 Scatter plot of scores from principal component scores 1 and 2, illustrating the PCA classification of normal and malignant tissue samples.

tation is useful because peaks in the loading correspond with Raman peak positions, recorded in wavenumbers (cm^{-1}), for chemical species that are more abundant in the tumor samples, and troughs correspond with entities more abundant in the normal samples. To provide some pointers as to the possible origin of the differences between malignant and normal tissue, Fig. 4(a) displays the loading from principal component 1, overlaid by the spectrum recorded for cytochrome C. While a number of the troughs in the loading coincide with some major cytochrome C peaks, including features at 1585 , 1352 , and 1240 cm^{-1} , there are also some notable absences, particularly a prominent feature at 1450 cm^{-1} . The favorable correlation analysis in Fig. 4(b) (Spearman rho correlation coefficient = -0.65 , $p < 0.001$) may suggest that excess porphyrin within the normal samples could explain the majority of the variation between the normal and malignant samples. Nevertheless, the peak discrepancies noted preclude this as a definitive conclusion.

The second principal component was analyzed in a similar fashion to PC 1. The loading from the second principal component, trace (a), is shown in Fig. 4(c) and compared with a reference Raman spectrum of DNA, trace (b). While some of the major bands in the DNA again exhibit a satisfactory correlation with peaks in the loading [Fig. 4(d), Spearman rho coefficient = 0.60 , $p < 0.001$], some DNA features are absent in the loading plot (see Sec. 4).¹⁸

Being a linear multivariate technique, principal component analysis is limited to a linear separation of the data, and for this reason the use of a nonlinear classification technique was explored. Random forest classification¹⁷ is an augmented form of decision tree analysis, where many decision tree models are constructed by randomly selecting subgroups of samples from a training set to construct models for each subgroup. Each decision tree then classifies each sample from an independent test set, and the final classification for this sample is decided by a weighted “majority vote” from all these decision trees. A randomly chosen training set was used to train the software to classify normal and malignant spectra. 40 independent spectra were randomly chosen to test this classification model. Random forest classification, with 28 variables tried at each split, was constructed on the training set and applied to the test set, resulting in a test set sensitivity of 90% (2/20 tumor samples

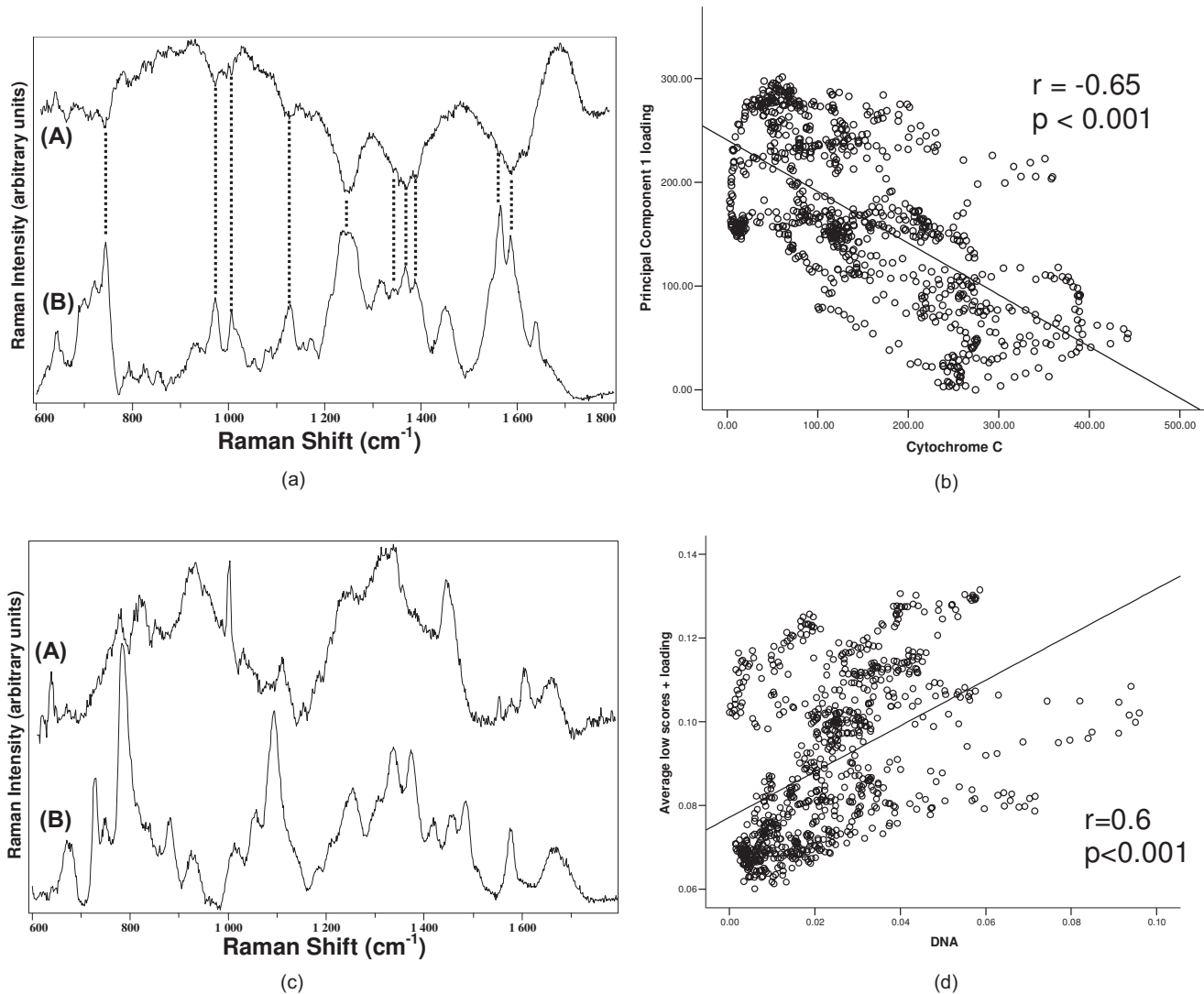


Fig. 4 (a) Illustration of (A) loading from first principal component from normal and malignant analysis overlaid by (B) reference Raman spectrum of cytochrome C. (b) Correlation scatter plot of loading from first principal component and the reference Raman spectrum of cytochrome C. (c) Illustration of (A) loading from principal component 2 and (B) the Raman spectrum of DNA. (d) Scatter plot of the loading from principal component 2 and the Raman spectrum of DNA, showing a positive correlation between these two spectra.

misclassified) and specificity of 75% (5/20 normal samples misclassified). The accuracy was unaffected by the number of variables tried (for values between 20 and 60).

Because Raman microscopy provides a unique chemical signature of tumor tissue, it provides a window into the chemical make-up of that tissue and may offer an insight into how aggressive a tumor is. Raman spectra from the tumor samples of 34 patients, for which there was at least 12 months postoperative follow-up, were analyzed using principal component analysis with leave-one-out cross-validation. The patients were divided into two groups depending on whether their cancer recurred within twelve months postoperatively. From this PCA, the first and third principal component scores separated the two groups. A scatter plot of these principal components is shown in Fig. 5. The diagonal line illustrates how these principal component scores can detect recurrence with a sensitivity of 73% (11/15) and a speci-

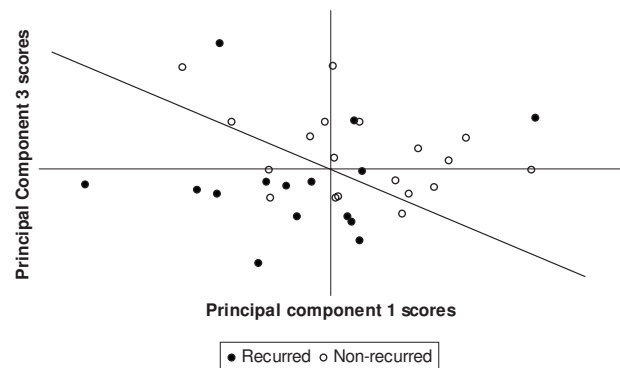


Fig. 5 Scatter plot of principal components 1 and 3 from the principal component analysis of patient recurrence spectra.

ficity of 74% (14/19), (positive predictive value 69% and negative predictive value 78%).

The tumor spectra were separated into two groups depending on whether the score from principal component 3 in the recurrence analysis was above or below its median. The overall median follow-up was 17 months. Using the log-rank test, there was no significant difference in survival between these two groups, although there was a strong trend ($p=0.086$).

4 Discussion

We have shown that Raman microscopy of normal and malignant tissue sections obtained from lung cancer resections can distinguish malignant and normal bronchial tissue and predict early postoperative recurrence in this group of patients.

Rather than relying on the use of individual spectral peaks to distinguish groups of tissue, multivariate data reduction techniques have been employed to include information from the whole spectrum to classify spectra. Principal component analysis (PCA) is a data reduction technique that summarizes the variation between the spectra by reducing it into a small number of principal component scores, while retaining the majority of information contained within the spectra. These components describe consistent differences between the spectra, allowing separation of the spectra into groups based on similarities or disparities in the spectral characteristics. This classification can be tested using a leave-one-out cross-validation, where the software develops a model based on all the spectra except one, and then tests the model on that spectrum, and repeats this process on all the spectra.

Principal component analysis of normal and malignant spectra demonstrated that Raman microscopy can distinguish normal and malignant lung tissue with a sensitivity of 84% and a specificity of 61%. Although the use of a nonlinear random forest classification technique improved the accuracy, the specificity was still suboptimal at 75%. This specificity, which is a function of the ability of a test procedure to classify normal samples, can be explained by the heterogeneous nature of the normal bronchial epithelium that was examined. Although all the normal samples were confirmed as “normal” by a pathologist, there were significant differences between these samples. It is likely that this morphological heterogeneity is also reflected in the cellular biochemistry and thus in the Raman spectra. The normal samples used in this study were taken from the same lung as the resected tumor and have certainly undergone the same inhaled insults as the tumor. They may also have undergone, to a varying degree, malignancy associated changes.^{19,20} It is also well established that tobacco smoking is a “field” effect, making the chemical analysis from this heterogeneous field challenging, although clinically relevant. As pointed out earlier, our classification accuracy is less than that reported in the two previous studies of lung cancer using Raman spectroscopy in the absence of the higher spatial resolution possible with a microscope.^{13,14} Some Raman scattering in the normal samples from these two previous studies will certainly have originated from stromal and supporting tissues, which may have unduly influenced the discrimination in these two studies. In our present study, only Raman scattering from normal and malignant cells was used, which can make the discrimination more challenging. It should be noted that in one of the earlier studies referred to,

only two Raman spectral peaks (1450 and 1662 cm^{-1}) were used to classify the samples. In our study, t-test analysis showed that there was no significant difference between the normal and tumor peak intensities at 1662 cm^{-1} , and multivariate analysis was required to exploit the subtle differences between the spectra, so as to classify the samples.

Two previous studies have shown that Raman spectroscopy can identify and grade prostatic adenocarcinoma tissue sections using the Gleason score as a gold standard, and identify chemoresistance in cell lines.^{9,21} However, this is the first study to suggest that Raman spectroscopy can provide prognostic information about lung cancer patients using clinical samples. A number of other methods have been used to predict prognosis in lung cancer, and a vast array of prognostic markers have been suggested, possibly reflecting the molecular heterogeneity of lung tumors.²² Recently, a number of studies have used gene expression signatures to divide patients undergoing surgical resection into a high- and low-risk group, depending on survival, and predict postoperative recurrence.^{23–27} Potti et al. applied a lung metagene model to independent cohorts from two multicenter studies and predicted recurrence with an accuracy of 72 and 79%, respectively. Larsen et al.²⁸ described a 54-gene signature that predicted recurrence in two independent cohorts with an accuracy of 72 and 67%. Our overall accuracy of 73% compares favorably with these studies. More recently, two groups have used a mass spectrometry proteomic approach to predict response to targeted chemotherapy and postoperative recurrence.²⁹ Raman spectroscopy inherently has two main advantages over other current technologies. Since many molecular and macromolecular species, including DNA, proteins and lipids, contribute to the Raman spectrum, the technique has the potential to tease out the molecular heterogeneity of lung tumors. Also, because Raman spectroscopy can be carried out under nondestructive conditions, there is the potential to obtain Raman spectra from lung tumors *in vivo* via bronchoscopy and thus obtain a diagnosis and prognosis simultaneously and noninvasively. Although this technology is still in development, *in vivo* fiber optic probes have been studied in lung cancer and other organ systems.^{6,7,30}

The analysis of the loading from the first PCA suggested that the main chemical difference between normal and malignant lung tissue may be attributable to a contribution of porphyrin to the spectrum. Although cytochrome C was used as the reference material, its Raman spectrum is virtually identical to that of the other cytochromes, because Raman scattering originates from the porphyrin ring structure of the heme component of the molecule. There are a number of intracellular cytochromes that this may represent, including cytochrome C, which is a key component in apoptosis.^{31,32} Evasion of this programmed cell death is characteristic of carcinogenesis. The tentative conclusion reached earlier (Sec. 3)—that cytochrome C may be more abundant in the normal bronchial epithelium than in the malignant lung cells—would be consistent with this.³³ A number of other cytochromes have been found in higher levels in bronchial epithelial cells when compared with lung cancer cells, including cytochrome oxidase, cytochrome P450, and CYP1A1.^{34–37} However, if there is indeed an excess of cytochrome (a cytoplasmic protein) in the normal cells, this could simply represent an increase in the

cytoplasm:nucleus ratio seen in normal tissue compared to malignant tissue. This is clearly an area that will require further investigation.

The analysis of the loading from principal component 2 in the first PCA suggested that this may be attributable to DNA, although some DNA peaks were absent in the loading trace. The conclusion though tentative would be consistent with the fact that DNA is more abundant in malignant tissue when compared to normal tissue because of the more active mitosis and cellular turnover.^{10,14,22}

There is clearly scope for further refinement of the methodology in future studies, such as increasing the number of spots for laser analysis in each sample map. Additionally, the use of fresh or thawed samples might have improved the accuracy of the discrimination, but this was logistically impossible in this current study.

In conclusion, Raman microscopy can differentiate malignant lung tissue from normal bronchial epithelium, and has the potential to predict early recurrence in patients undergoing lung cancer resection. However, this should be tested in a larger independent test set.

Acknowledgments

Research and Development Office for Health and Personal Social Services in Northern Ireland (EAT/2538/03); and The Biotechnology and Biological Sciences Research Council (18471). Neither funding agency has had a role in the design of the study; collection, analysis, or interpretation of the data; the decision to submit the manuscript for publication; or the writing of the manuscript.

References

- J. Ferlay, P. Autier, M. Boniol, M. Heanue, M. Colombet, and P. Boyle, "Estimates of the cancer incidence and mortality in Europe in 2006," *Ann. Oncol.* **18**, 581–592 (2007).
- J. Hung, S. Lam, J. C. LeRiche, and B. Palcic, "Autofluorescence of normal and malignant bronchial tissue," *Lasers Surg. Med.* **11**, 99–105 (1991).
- A. M. McWilliams, J. R. Mayo, M. I. Ahn, S. L. MacDonald, and S. C. Lam, "Lung cancer screening using multi-slice thin-section computed tomography and autofluorescence bronchoscopy," *J. Thorac. Oncol.* **1**, 61–68 (2006).
- C. V. Raman and K. S. Krishnan, "A new type of secondary radiation," *Nature* **121**, 501–502 (1928).
- R. L. McCreery, *Raman spectroscopy for chemical analysis*, Wiley-Interscience, New York (2000).
- H. Zeng, J. Zhao, M. Short, D. I. McLean, S. Lam, A. McWilliams, and H. Lui, "Raman spectroscopy for *in vivo* tissue analysis and diagnosis, from instrument development to clinical applications," *J. Inno. Health Sci.* **1**(1), 95–106 (2008).
- M. A. Short, S. Lam, A. McWilliams, J. Zhao, H. Lui, and H. Zeng, "Development and preliminary results of an endoscopic Raman probe for potential *in vivo* diagnosis of lung cancers," *Opt. Lett.* **33**(7), 711–713 (2008).
- M. V. Chowdary, K. K. Kumar, J. Kurien, S. Mathew, and C. M. Krishna, "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy," *Biopolymers* **83**(5), 556–569 (2006).
- P. Crow, N. Stone, C. A. Kendall, J. S. Uff, J. A. Farmer, H. Barr, and M. P. Wright, "The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*," *Br. J. Cancer* **89**(1), 106–108 (2003).
- A. Mahadevan-Jansen, M. F. Mitchell, N. Ramanujam, A. Malpica, S. Thomsen, U. Utzinger, and R. Richards-Kortum, "Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers," *Photochem. Photobiol.* **68**, 123–132 (1998).
- N. Stone, C. Kendall, J. Smith, P. Crow, and H. Barr, "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss.* **126**, 141–157 (2004).
- C. J. Frank, R. L. McCreery, and D. C. Redd, "Raman spectroscopy of normal and diseased human breast tissues," *Anal. Chem.* **67**, 777–783 (1995).
- Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, and H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer* **107**(6), 1047–1052 (2003).
- H. Yamazaki, S. Kaminaka, E. Kohda, M. Mukai, and H. O. Hamaguchi, "The diagnosis of lung cancer using 1064-nm excited near-infrared multichannel Raman spectroscopy," *Radiat. Med.* **21**, 1–6 (2003).
- S. Koljenovic, T. C. Bakker Schut, J. P. van Meerbeeck, A. P. Maat, S. A. Burgers, P. E. Zondervan, J. M. Kros, and G. J. Puppels, "Raman microspectroscopic mapping studies of human bronchial tissue," *J. Biomed. Opt.* **9**, 1187–1197 (2004).
- A. M. C. Davies and T. Fearn, "Back to basics: the principles of principal component analysis," *Spectroscopy Europe* **16**, 20–23 (2004).
- L. Breiman, "Random forests," *Mach. Learn.* **45**, 5–32 (2001).
- W. A. Franklin, "Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia," *Chest* **117**, 80S–89S (2000).
- N. Ikeda, C. MacAulay, S. Lam, J. LeRiche, P. Payne, D. Garner, C. Konaka, H. Kato, and B. Palcic, "Malignancy associated changes in bronchial epithelial cells and clinical application as a biomarker," *Lung Cancer* **19**, 161–166 (1998).
- T. B. van Oppen, A. Meisels, and S. St. Pierre, "Malignancy associated changes in sputum: a correlated study of 315 patients," *Acta Cytol.* **19**, 573–576 (1975).
- C. M. Krishna, G. Kegelaer, I. Adt, S. Rubin, V. B. Kartha, M. Manfait, and G. D. Sockalingum, "Combined Fourier transform infrared and Raman spectroscopic approach for identification of multidrug resistance phenotype in cancer cell lines," *Biopolymers* **82**(5), 462–470 (2006).
- M. D. Brundage, D. Davies, and W. J. Mackillop, "Prognostic factors in non-small cell lung cancer: a decade of progress," *Chest* **122**, 1037–1057 (2002).
- A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson, "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13790–13795 (2001).
- M. E. Garber, O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, R. M. van de, G. D. Rosen, C. M. Perou, R. I. Whyte, R. B. Altman, P. O. Brown, D. Botstein, and I. Petersen, "Diversity of gene expression in adenocarcinoma of the lung," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13784–13789 (2001).
- M. Raponi, Y. Zhang, J. Yu, G. Chen, G. Lee, J. M. Taylor, J. MacDonald, D. Thomas, C. Moskaluk, Y. Wang, and D. G. Beer, "Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung," *Cancer Res.* **66**, 7466–7472 (2006).
- Y. Lu, W. Lemon, P. Y. Liu, Y. Yi, C. Morrison, P. Yang, Z. Sun, J. Szoke, W. L. Gerald, M. Watson, R. Govindan, and M. You, "A gene expression signature predicts survival of patients with stage I non-small cell lung cancer," *PLoS Med.* **3**, e467 (2006).
- A. Potti, S. Mukherjee, R. Petersen, H. K. Dressman, A. Bild, J. Koontz, R. Kratzke, M. A. Watson, M. Kelley, G. S. Ginsburg, M. West, D. H. Harpole, Jr., and J. R. Nevins, "A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer," *N. Engl. J. Med.* **355**, 570–580 (2006).
- J. E. Larsen, S. J. Pavey, L. M. Passmore, R. V. Bowman, N. K. Hayward, and K. M. Fong, "Gene expression signature predicts recurrence in lung adenocarcinoma," *Clin. Cancer Res.* **15**, 2946–2954 (2007).
- F. Taguchi, B. Solomon, V. Gregorc, H. Roder, R. Gray, K. Kasahara, M. Nishio, J. Brahmer, A. Spreafico, V. Ludovini, P. P. Massion, R. Dziadziuszko, J. Schiller, J. Grigorieva, M. Tsybin, S. W. Hunsucker, R. Caprioli, M. W. Duncan, F. R. Hirsch, P. A. Bunn, Jr., and D. P. Carbone, "Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study," *J. Natl. Cancer Inst.* **99**, 838–846 (2007).

30. P. Matousek, E. R. Draper, A. E. Goodship, I. P. Clark, K. L. Ronayne, and A. W. Parker, "Noninvasive Raman spectroscopy of human tissue *in vivo*," *Appl. Spectrosc.* **60**, 758–763 (2006).
31. X. Liu, C. N. Kim, J. Yang, R. Jemmerson, and X. Wang, "Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome C," *Cell* **86**, 147–157 (1996).
32. X. Saelens, "Toxic proteins released from mitochondria in cell death," *Oncogene* **23**, 2861–2874 (2004).
33. D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell* **100**, 57–70 (2000).
34. M. J. Hannibal, M. M. Nachlas, and A. M. Seligman, "Histochemical localization of cytochrome oxidase in human tissues and neoplasms," *Cancer* **13**, 1008–1016 (1960).
35. E. T. Cantrell, G. A. Warr, D. L. Busbee, and R. R. Martin, "Induction of aryl hydrocarbon hydroxylase in human pulmonary alveolar macrophages by cigarette smoking," *J. Clin. Invest.* **52**, 1881–1884 (1973).
36. S. Petruzzelli, A. M. Camus, L. Carrozzi, L. Ghelarducci, M. Rindi, G. Menconi, C. A. Angeletti, M. Ahotupa, E. Hietanen, and A. Aitio, "Long-lasting effects of tobacco smoking on pulmonary drug-metabolizing enzymes: a case-control study on lung cancer patients," *Cancer Res.* **48**, 4695–4700 (1988).
37. T. L. McLemore, S. Adelberg, M. C. Liu, N. A. McMahon, S. J. Yu, W. C. Hubbard, M. Czerwinski, T. G. Wood, R. Storeng, and R. A. Lubet, "Expression of CYP1A1 gene in patients with lung cancer: evidence for cigarette smoke-induced gene expression in normal lung tissue and for altered gene regulation in primary pulmonary carcinomas," *J. Natl. Cancer Inst.* **82**, 1333–1339 (1990).