Real-time endoscopic Raman spectroscopy for \textit{in vivo} early lung cancer detection

Hanna C. McGregor\textsuperscript{1, 2}, Michael A. Short\textsuperscript{1}, Annette McWilliams\textsuperscript{1}, Tawimas Shaipanich\textsuperscript{1, 4}, Diana N. Ionescu\textsuperscript{3}, Jianhua Zhao\textsuperscript{1}, Wenbo Wang\textsuperscript{1}, Guannan Chen\textsuperscript{1}, Stephen Lam\textsuperscript{1, 2, 4}, and Haishan Zeng\textsuperscript{*}, 1, 2, 3

\textsuperscript{1} Imaging Unit – Integrative Oncology Department, BC Cancer Agency Research Centre, Vancouver, British Columbia, Canada
\textsuperscript{2} Interdisciplinary Oncology Program, University of British Columbia, Vancouver, British Columbia, Canada
\textsuperscript{3} Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada
\textsuperscript{4} Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

Received 30 July 2015, revised 22 October 2015, accepted 24 November 2015
Published online 9 January 2016

**Key words:** \textit{in vivo} tissue Raman spectroscopy, endoscopy, lung cancer, early cancer detection, spectral diagnosis

Currently the most sensitive method for localizing lung cancers in central airways is autofluorescence bronchoscopy (AFB) in combination with white light bronchoscopy (WLB). The diagnostic accuracy of WLB + AFB for high grade dysplasia (HGD) and carcinoma \textit{in situ} is variable depending on physician’s experience. When WLB + AFB are operated at high diagnostic sensitivity, the associated diagnostic specificity is low. Raman spectroscopy probes molecular vibrations and gives highly specific, fingerprint-like spectral features and has high accuracy for tissue pathology classification. In this study we present the use of a real-time endoscopy Raman spectroscopy system to improve the specificity. A spectrum is acquired within 1 second and clinical data are obtained from 280 tissue sites (72 HGDs/malignant lesions, 208 benign lesions/normal sites) in 80 patients. Using multivariate analyses and waveband selection methods on the Raman spectra, we have demonstrated that HGD and malignant lung lesions can be detected with high sensitivity (90%) and good specificity (65%).

1. Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide with an overall 5 year survival rate of just 17% after diagnoses \cite{1}. The reasons for the poor prognosis are that patients tend to be diagnosed at an advanced stage, coupled with a lack of effective treatments for them. This is in contrast to
patients found with early stage (0) carcinoma in situ (CIS) or stage 1A (tumor less than 2 cm without metastatic spread) where the 5 year survival is >70% [2]. Bronchoscopy has been used for decades for localizing early stage cancers of the central airways, during which time there have been many technological improvements to the bronchoscope itself, and the development of helpful adjunct devices [3–9]. In the 1990s autofluorescence bronchoscopy (AFB) was developed and increasingly used in adjunct to the standard white light bronchoscopy (WLB) to improve the localization of early stage cancers and high grade dysplasia (HGD: moderate dysplasia and severe dysplasia) of the main airways [3, 4]. A recent meta-analyses of 21 studies performed by Sun et al. [10] showed that combined WLB + AFB exam greatly improved the sensitivity for localizing HGD or CIS, but with decreased specificity as compared to WLB exam alone. The pool relative sensitivity of WLB + AFB versus WLB was 2.04, while the pool relative specificity of WLB + AFB versus WLB was 0.65. This highlights the need of developing new adjunct technologies for improving the specificity. Although the evidence for progression of a lesion with HGD to invasive cancer (IC) is weak, there is strong evidence for a field effect where HGD is a risk factor for the development of IC somewhere in the lung [11]. Thus identifying lesions in the central airways with HGD provides valuable information about patients requiring close follow-up. There was however considerable variation in the sensitivities (41–100%) obtained across all 21 studies that used a combined WLB + AFB for indentifying HGD. The largest variability in sensitivities occurred at high specificities which probably reflect differences in biopsy protocols, and operator experience. Those data suggested that in order to increase the probability of achieving a high sensitivity a liberal biopsy protocol is required otherwise the risk of missing a HGD escalates significantly. The drawback with such an approach is that it will result in many biopsies which are negative for HGD or worse, which may lead to longer procedural times, greater healthcare costs and a higher incidence of procedure complications. These factors hinder the practical adoption of this technology for widespread clinical uses. The same meta-analysis also showed that the specificity obtained across all 21 studies varied from 18–86%. The general trends are that for studies achieved high sensitivities with WLB + AFB, their specificities were relatively low.

It is desirable for a clinical diagnostic tool to have high diagnostic sensitivity and good specificity. We carefully reviewed all the studies cited in the meta-analysis and focused on studies differentiating HGD + CIS from benign lesions with large number of cohorts and good prevalence. Three such studies [12–14] suggested that it was possible to achieve better than 90% sensitivity for HGD + CIS localization using WLB + AFB although the associated specificity became low (between 18% and 32%). Our goal is to develop a new adjunct technology, such as point Raman spectroscopy that is capable of providing high specificity for HGD/CIS identification, to be used together with a WLB + AFB system operating at high sensitivity level (>90%) for improved HGD/CIS localization. This will ultimately lead to practical early detection of lung cancers, thus greatly improved survival rate.

A number of technologies have been explored to improve the specificity of an AFB + WLB examination. The ratio of red to green (R/G) fluorescence of a lesion is the easiest to implement requiring minimal additional equipment, but so far the results have been mixed. Another technology that has been explored by our group and others to improve the specificity in identifying HGD during a bronchoscopy is reflectance spectroscopy. However, initial in vivo tests were conducted on a relatively small number of lesions which had either a benign/normal or IC pathology; very few, if any HGD were included [15, 16]. Nevertheless the concept should be explored further especially with our latest technology development that does not require a separate optical fibre catheter, but can obtain the reflectance spectra directly from the WLB imaging camera [16].

In contrast to the broad spectral features of fluorescence and reflectance technology, Raman spectroscopy, based on inelastic light scattering, probes molecular vibrations and gives very specific, fingerprint-like spectral features and has high accuracy for differentiation between malignant and benign tissues [17]. It is potentially useful for improving the specificity of early lung cancer localization. However, the Raman signal is exceedingly low, and as a consequence, long integration times are required to acquire sufficiently strong Raman signals for a single spectrum. A traditional Fourier-transform Raman system requires up to 30 minutes of integration time to acquire one Raman spectrum with reasonably good signal-to-noise ratio (SNR). This has hindered the clinical applications of Raman technology. Our group has successfully developed a rapid, real-time Raman spectrometer system and a dedicated endoscopy Raman catheter for lung measurements that substantially reduces the spectral acquisition time to less than 1 second [18]. This system employed proprietary technologies for improving the spectrometer’s SNR [19], Raman catheter fiber background fluorescence elimination, and size miniaturization [18]. We realized real-time in vivo Raman measurements of the lung for the first time [18]. In a pilot study, we conducted point Raman spectroscopy measurements of suspicious areas identified by WLB + AFB imaging for improving early lung cancer detection [20]. The results very well demon-
strated the technical feasibility and clinical compatibility. In this manuscript, we report a study using the developed Raman spectroscopy technology as an adjunct device to WLB + AFB exam to improve the specificity of localizing HGD/CIS of the central lung airways, while maintaining high detection sensitivity.

2. Patients and methods

This study was approved by the University of British Columbia – BC Cancer Agency Research Ethics Board (certificate number: H06–00010). Patients who were attending the BC Cancer Agency Vancouver Center for a previously scheduled bronchoscopy were invited to volunteer for this study. Patients must have already consented to a bronchoscopy as part of a standard diagnostic procedure or as part of an approved lung cancer prevention study, before being approached to volunteer. Patients were excluded if they had a cardiac pacemaker or implanted defibrillator device, had a known allergic reaction to Xylocaine, were taking a blood thinner such as warfarin or heparin, or had any medical condition such as acute or chronic respiratory failure, which could jeopardize the safety of the patient during participation in the study. Women who were premenopausal were excluded unless they were surgically sterile or on the birth control pill.

The Raman system used to take measurements was similar to the ones described previously by our group [18, 20]. The main differences were the inclusion of a new thermal electrically cooled CCD detector for faster start up times and reduced optical noises (etaloning effect), as well as a new spectrograph with a holographic reflection type grating, allowing tunable wavelength range [21]. These changes were implemented to improve the SNR, thus allowing for a more reliable extraction of the Raman signal from the fluorescence background instead of the more obtuse and glossy 2nd order derivative processing of the data used by us previously [20]. Figure 1 shows the schematic diagram of the endoscopic laser Raman spectroscopy system. The inserts show the arrangement of the excitation (red) and collection fibers (green). The Raman excitation light was produced by a wavelength stabilized 785 nm diode laser, and delivered to the tissue surface by a detachable 1.8 mm size fiber optic probe (Raman catheter) passed down the instrument channel of the bronchoscope. The maximum excitation power at the tissue surface was 150 mW. The same catheter collected emission from the tissue and delivered it to the spectrometer for analyses. The collection fibers were connected to the spectrophotograph through a special round to parabolic fiber bundle to correct the spectral imaging distortion to achieve better SNR.

OH impurity fibers, and a gold coated excitation fiber to avoid cross-talk between the excitation and collection fibers. Optical filters were coated at the distal end of the probe to filter out laser noise, fiber emission, and to attenuate all collected light with wavelengths \( \leq 820 \) nm (\( \leq 540 \) cm\(^{-1} \) relative to the 785 nm excitation). At its proximal end, the probe was attached to a second set of optical filters with similar transmission characteristics, but higher OD (optical density) in the rejection bands for further, and better, attenuation of the aforementioned unwanted emissions. The spectrograph grating was tuned to cover the high wavenumber range of 2050–3100 cm\(^{-1} \) which is known to have much less tissue autofluorescence than lower wavenumbers (\(<1800\) cm\(^{-1} \)), and yet still have Raman bands sensitive to biomolecular changes [18, 20]. The spectral resolution in this wavenumber range was estimated to be 8 cm\(^{-1} \). Although the SNR increases with excitation energy (power \( \times \) time), an acquisition time of 1 second was used as this was the time that the catheter and excitation spot could be reliably maintained in the same position on the tissue surface. After data acquisition, custom designed software was used to subtract the fluorescence background using a modified polynomial fit (9th order) and to display the calibrated Raman spectrum all within a fraction of a second [22]. An example of the fluorescence background removal was shown in Supplementary Information Figure S1.

The procedure was for patients to first undergo a standard WLB + AFB exam which took place in the Endoscopy Suite and the location of any autofluor-
escence positive lesions identified. A three stage WLB + AFB visual grading of each site was determined by the physician following the criteria listed in Table 1 [4]. After the AFB, the Raman probe was inserted into the instrument channel of the bronchoscope by the physician and directed toward an auto-fluorescence positive lesion or a normal tissue control site. A probe tip to tissue distance of between 5–10 mm was used which generated an excitation illumination spot diameter on the tissue surface of between 2–4 mm. A lesion was measured between 1–6 times, depending on size, and different lesions were considered to be separate sites with unique distinct pathology for analyses. Biopsies were taken as directed by the physician, and analyzed by an experienced pathologist using standard histopathology assessment for lung lesions [4]. The pathological coding system and the corresponding tissue diagnosis are listed in Table 2. If different biopsy fragments of the same tissue site had different histological assessments, the worst assessment was taken for analysis. Multiple Raman spectra from a tissue site were averaged to generate a single spectrum to represent the site. The final data file matched the histopathology results to the WLB + AFB visual grades and the Raman spectrum for each site if all possible.

Between May 2011 and November 2012, Raman spectra were obtained from 80 patients. Raman measurements were taken from a total of 280 sites. Of the 280 sites, 214 sites were biopsied after the Raman measurements and histologically assessed. Pathology reports showed that 72 sites were graded as either a HGD or worse, 90 sites were graded as benign (inflammation, metaplasia, hyperplasia or mild dysplasia), and 52 were graded normal. Although biopsying normal tissue was not standard clinical procedure, some normal control sites were biopsied as these were part of another study with a protocol which required the subject to consent to the taking of additional tissue. The remaining 66 control sites were not biopsied; instead they were determined to be normal by the physician through visual grading during the WLB and AFB exam. Detailed patient demographics are shown in Table 3.

### Table 1 Visual grading system and the corresponding tissue description.

<table>
<thead>
<tr>
<th>Visual Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Normal)</td>
<td>No visual abnormality</td>
</tr>
<tr>
<td>2 (Abnormal)</td>
<td>Visual changes suggestive of inflammation, trauma, hyperplasia, metaplasia, mild dysplasia</td>
</tr>
<tr>
<td>3 (Suspicious)</td>
<td>Visual changes suggestive of moderate dysplasia, severe dysplasia, carcinoma \textit{in situ} or invasive cancer</td>
</tr>
</tbody>
</table>

### Table 2 Pathological coding system and the corresponding tissue diagnosis.

<table>
<thead>
<tr>
<th>Pathological Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal &amp; Benign</td>
<td>1 Normal</td>
</tr>
<tr>
<td></td>
<td>2 Inflammation</td>
</tr>
<tr>
<td></td>
<td>3 Hyperplasia or Metaplasia</td>
</tr>
<tr>
<td></td>
<td>4 Mild Dysplasia</td>
</tr>
<tr>
<td>HGD* &amp; Malignant</td>
<td>5 Moderate or Severe Dysplasia</td>
</tr>
<tr>
<td></td>
<td>6 Carcinoma \textit{in situ} (CIS)</td>
</tr>
<tr>
<td></td>
<td>7 Microinvasive Cancer</td>
</tr>
<tr>
<td></td>
<td>8 Invasive Cancer (IC)</td>
</tr>
</tbody>
</table>

* HGD refers to High Grade Dysplasias, specifically Moderate Dysplasia and Severe Dysplasia

### Table 3 Patient demographics and location of the Raman reading. A total of 280 sites were measured.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Patient Demographics</th>
<th>Raman Reading Location</th>
<th>Central (trachea)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive SCC and CIS</td>
<td>Mean age: 66 (41–86)</td>
<td>Male: 21</td>
<td>Female: 10</td>
</tr>
<tr>
<td>Severe + Moderate Dysplasia</td>
<td>Mean age: 62 (40–78)</td>
<td>Male: 32</td>
<td>Female: 9</td>
</tr>
<tr>
<td>Mild Dysplasia + Metaplasia</td>
<td>Mean age: 62 (40–80)</td>
<td>Male: 52</td>
<td>Female: 16</td>
</tr>
<tr>
<td>Hyperplasia</td>
<td>Mean age: 67 (40–78)</td>
<td>Male: 14</td>
<td>Female: 2</td>
</tr>
<tr>
<td>Inflammation</td>
<td>Mean age: 65 (61–72)</td>
<td>Male: 6</td>
<td>Female: 0</td>
</tr>
<tr>
<td>Normal</td>
<td>Mean age: 67 (40–84)</td>
<td>Male: 86</td>
<td>Female: 32</td>
</tr>
<tr>
<td>Total</td>
<td>Mean age: 211 (40–84)</td>
<td>Male: 69</td>
<td>Female: 69</td>
</tr>
</tbody>
</table>
3. Statistical analysis

Multivariate statistical methods have been widely used for classification of Raman spectra in a number of studies related to cancer diagnoses including skin and were proven to be very effective and reliable [17]. In this study, principal components with generalized discriminant analysis (PC-GDA) and partial least-squares (PLS or called PLS-DA) were used for spectral classification with leave-one-out cross-validation (LOO-CV) where successive single spectra were left out for test with the remaining spectra used for training.

For PC-GDA analysis, we followed the methodology in [23]. Briefly, the training process consisted of the following procedures: (1) the mean and standard deviation of the training spectra data set were calculated, (2) each spectrum in the training data set was standardized by subtracting the mean and then dividing by the standard deviation, (3) the standardized training spectra were analyzed with principal component analysis. PC factors of the training cases and PC loadings were obtained, and (4) a generalized (linear) discrimination model was developed from the PC factors which could be used directly to predict the new cases. The testing process consisted of the following procedures: (1) the test spectrum was standardized by removing the mean and dividing by the standard deviation obtained from the training spectra, (2) the PC factors of the test spectrum was calculated based on the PC loadings from the training spectra, and (3) a posterior probability of the testing spectrum was obtained from the discrimination model developed in the training procedure. The above procedures were repeated until all the spectra were left out once (and only once) for testing.

For PLS analysis, we followed the methodology in [24]. The training process consisted of the following procedures: (1) the mean and standard deviation of the training spectra data set were calculated, (2) each spectrum in the training data set was standardized by subtracting the mean and then dividing by the standard deviation, (3) the standardized training spectra were analyzed with NIPALS algorithm (non-linear iterative partial least squares) with the classification of the training spectra setting to the known value. The weight factors, the loadings, the regression coefficient and the factor scores of the training spectra were obtained, (4) a general discrimination model was developed from the training spectra which could be used directly to predict new cases. The testing process consisted of the following procedures: (1) the test spectrum was standardized by removing the mean and dividing by the standard deviation obtained from the training spectra, (2) the factor scores of the test spectrum were calculated based on the weight factors from the training spectra, and (3) a posterior probability of the testing spectrum was obtained from the discrimination model developed in the training procedure. The above procedures were repeated until all the spectra were left out once (and only once) for testing.

Before the statistical analysis, all the Raman spectra were normalized to their respective integrated spectral areas under the curve (AUC). PCA-DA and PLS analyses were performed on the full spectra based data set and also on data set with selected discrete wavebands generated using a number of feature selection strategies including: stepwise multiple regression (STEP) [25], least absolute shrinkage selection operator (LASSO) [26], and genetic algorithm (GA) [27]. These feature selection algorithms have been previously utilized to improve model prediction/classification in a number of studies, for example, genomics and proteomics [28], lung cancer [29], breast cancer [30], and oral cancers [31]. All the multivariate classification analyses in this study were implemented using MATLAB (version 2013b, Math-Works).

Waveband selection results depend on the sample spectra and sample size. In order to find reliable optimal wavebands, a LOO-CV protocol was used. In the LOO-CV waveband selection procedure, a single spectrum was left out with the remaining spectra used for waveband selection operation. A set of wavebands were selected which gave the best diagnostic performance of the training spectra. By repeating this procedure, every spectrum was left out once for wavelength selection purpose. At the end, n sets of wavebands were selected where n was the total number of cases. The n sets of wavebands were then accumulated. The wavebands with higher odds from the LOO-CV analysis were chosen for subsequent PCA-DA and PLS analyses. Waveband selection was also tested using three-fold cross-validation where the spectra were divided equally and randomly into three groups with two groups for training and one group for testing. Similar results were obtained for three-fold cross-validation with those of LOO-CV. In order to prevent over-fitting or selecting spurious wavebands, a window size of 5 pixels, corresponding to an average of 5.3 cm⁻¹ spectral range, was chosen based on recommendations from Refs. [32, 33].

The receiver operating characteristic (ROC) curve was calculated from the posterior probabilities derived from each of the analysis models described above and represents the diagnostic performance of each model. The AUC of each ROC was calculated using the trapezoidal rule [34]. The significance of these AUCs and comparisons between different AUCs were carried out in a standard fashion [35]. All ROC analyses were based on nonparametric techniques and were conducted separately for the PC-GDA and PLS analyses. To compare the differ-
ent statistical methods used, and to compare the utility of Raman spectroscopy with other non-invasive diagnostic techniques, the specificities were calculated for fixed sensitivity levels of: 90% and 95%.

4. Results

The Raman spectra were dominated by strong CH stretching related bands in the wavenumber range from 2775 cm⁻¹ to 3040 cm⁻¹. For wavenumbers below 2775 cm⁻¹ the spectra did not contain any clear Raman signals, but a significant amount of noise attributed to etaloning effects and residual optical fibre emissions. The mean Raman spectra from 2775 cm⁻¹ to 3040 cm⁻¹ for each histopathology group are shown in Figure 2. Mild dysplasia and metaplasia were grouped together because both pathologies are considered to be low grade preneoplastic lesions, carrying the same risk of progression. Similarly, moderate and severe dysplasias were grouped together because both are considered to be high grade preneoplastic lesions which carry a similar risk of progression into invasive cancer. Furthermore, due to a small number of CIS lesions (n=2) as well as the higher risk of progression into invasive cancer, CIS was grouped with invasive cancers (IC). And we did not get any microinvasive cancer cases in this study. For these reasons, Figure 2 only has 6 categories. Major Raman peaks are seen at 2850 cm⁻¹, 2885 cm⁻¹, 2940 cm⁻¹, 2965 cm⁻¹, 2990 cm⁻¹, and 3020 cm⁻¹. These peaks were assigned to various fundamental CH, CH₂, and CH₃ stretching modes [36] and overtones of CH₂ and CH₃ bending modes [37]. The peak at 2850 was assigned to the CH₂ symmetric stretching modes of fatty acids and lipids, while the peak at 2885 cm⁻¹ was for the CH₃ symmetric stretching modes. [36]. The main peak at 2940 cm⁻¹ was assigned to a mixture of CH vibrations in proteins and CH₃ asymmetric stretching modes of lipids and nucleic acids [20, 36]. And the peaks at 2965 cm⁻¹ and 2990 cm⁻¹ were assigned to in-plane and out-of-plane anti-symmetric CH₃ stretching in lipid and fatty acid molecules [36, 38, 39]. The peak 3020 cm⁻¹ was assigned to the asymmetric stretching of =C–H group in RCH=CHR molecules [39], in which the R stands for an alkene functional group.

Apart from the main Raman peaks there was evidence for smaller Raman peaks, or inflection points at 2790 cm⁻¹, 2825 cm⁻¹ and 2920 cm⁻¹ that did not appear to be related to noise. The origins of these bands were uncertain. The 2920 cm⁻¹ band was most likely due to Fermi resonance interactions between the main stretching modes and CH bending overtones. Whereas the 2825 cm⁻¹ band may be due to one of the pair of CH stretching modes of aldehydic functional groups, with the other lost in the noise at lower wavenumbers [40]. No explanation can be offered currently for the 2790 cm⁻¹ band.

Despite there being clear Raman bands that were probably connected to the abundance of different biomolecules, there were no unique peaks that could be assigned to lung cancer alone. Although on average there was a distinctive loss of the lipid peak at 2850 cm⁻¹ seen in the spectra from malignant lesions, the amount lost for individual lesions showed considerable variation. There were also differences in the spectra from inflamed tissue compared to the other pathologies. Specifically, intensity of the inflammation group was relatively higher than all other categories between 2850 cm⁻¹ and 2900 cm⁻¹. To extract a more reliable correlation of spectra with pathology, multivariate statistical techniques were used.

**Figure 2** Mean Raman spectra by diagnosis. All spectra were normalized to their respective area under curve (AUC) before averaging by diagnosis. CIS: carcinoma in situ.
4.1 Classification based on HGD and malignant versus benign and normal lung tissues

When the Raman spectra (full range) were used to distinguish HGD and malignant (CIS & invasive cancers) lung tissues \((n = 72)\) from benign lung diseases and normal lung tissues \((n = 208)\), the AUCs of the resulting ROCs for PLS and PC-GDA analyses were almost identical at 0.83 (column 2, Table 4) and statistically significant \((p < 0.001)\). Waveband selection techniques STEP, LASSO and GA were applied before PC-GDA and PLS analysis to improve the diagnostic performances. Figure S2 and Figure S3 in the Supplementary Information show the wavebands selected by STEP and LASSO respectively. Some of the selected wavebands, but not all of them, were located at the Raman peaks. Some regions off the peak positions of the spectra were also selected and able to help with classification as well. The three waveband selection methods increased the AUC of their respective ROCs to between 0.85 and 0.88 (columns 3–5, Table 4). For a 90% (95% CI: 0.81–0.96) sensitivity the PLS analyses with STEP waveband section provided the best specificity of 51% (95% CI: 0.44–0.58). As an example, the posterior probabilities and the ROC curve corresponding to the STEP PC-GDA analysis are shown in Figure 3.

Figure 3A shows the posterior probability (PP) for each lesion to be classified as a lung cancer (malignant) or precancerous lesion (HGD). From the distribution of the posterior probabilities, the ROC curve with 95% CIs is generated and shown in Figure 3B. Figure 3C shows the box plot representation of the posterior probability distributions according to lesion subcategories. The PP values are relatively low for normal tissue and all benign lesion categories, but increase greatly for the HGD and malignant groups.

4.2 Classification based on modeling malignant versus normal lung tissues

A different classification of all the histopathologies \((n = 72 \text{ vs. } n = 208)\) was tried by using models generated from spectra of the two extremes only: IC \((n = 29)\) and normal sites \((n = 118)\). These models were then used to classify all histopathologies. The rationale for this was that all histopathologies were expected to lie between IC and normal sites according to different levels of sensitivity for full range and waveband selection algorithms. HGD and malignant \((n = 72)\) versus benign lung lesions and normal lung tissues \((n = 208)\).

<table>
<thead>
<tr>
<th>Sensitivity level (95% CI)</th>
<th>Full range</th>
<th>STEP</th>
<th>LASSO</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95 (0.88–0.99)</td>
<td>0.42</td>
<td>0.49</td>
<td>0.46</td>
<td>0.44</td>
</tr>
<tr>
<td>0.90 (0.81–0.96)</td>
<td>0.51</td>
<td>0.65</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>0.95 (0.88–0.99)</td>
<td>0.40</td>
<td>0.46</td>
<td>0.43</td>
<td>0.44</td>
</tr>
<tr>
<td>0.90 (0.81–0.96)</td>
<td>0.51</td>
<td>0.64</td>
<td>0.46</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 4 Area under the ROC curve and 95% CI based on full range and waveband selection algorithms for discriminating precancers (HGD) and malignant lung cancers \((n = 72)\) from benign lung diseases and normal lung tissues \((n = 208)\).

Table 5 Summary of Raman spectroscopy diagnostic parameters derived from ROCs: specificity values according to different levels of sensitivity for full range and waveband selection algorithms. HGD and malignant \((n = 72)\) versus benign lung lesions and normal lung tissues \((n = 208)\).
These tests provided an ROC AUC of 0.85 (95% CI: 0.81–0.90), based on full spectrum PC-GDA, and an ROC AUC of 0.86 (95% CI: 0.81–0.91) based on full spectrum PLS. The posterior probabilities and the ROC curve corresponding to the full spectrum PC-GDA analysis are shown in Figure 4, and demonstrated that all the histopathologies, including those categories not trained, can be well classified based on this approach.

Figure 3  Lesion classification by Raman spectroscopy based on STEP PC-GDA analysis. (A) The posterior probability plot for distinguishing cancerous lesions (HGD, CIS and invasive cancer; \(n = 72\)) from benign lesions (mild dysplasia, metaplasia, hyperplasia, inflammation and normal; \(n = 208\)) (B) The ROC curves and 95% CIs derived from the posterior probabilities. (C) The box plot representation of the posterior probability distributions according to lesion subcategories. The lower bound on each box shows the 25th percentile, where the upper bound shows the 75th percentile, meaning 25% and 75% of the data points are below these two bounds respectively. The line between these two bounds is the median. The whiskers on the plot show the 5th and 95th percentile, meaning 5% and 95% of the data are below these whiskers respectively. Data found outside the 5–95th percentile whiskers are outliers, as shown by separate data points.

Figure 4  Lesion classification based on modeling using the extremes cases (29 Invasive Cancers and 118 Normal Sites) for training (full spectra PC-GDA). Testing was done for the entire data set (72 HGD/malignant vs. 208 benign/normal). (A) The posterior probability plot for distinguishing cancerous lesions (HGD, CIS and invasive cancer; \(n = 72\)) from benign lesions (mild dysplasia, metaplasia, hyperplasia, inflammation and normal; \(n = 208\)). (B) The ROC curves and 95% CIs derived from the posterior probabilities. (C) The box plot representation of the posterior probability distributions according to lesion subcategories.
4.3 Sensitivity and specificity of visual diagnosis

Before performing Raman spectrum measurement, a visual grade was also determined for the majority of sites during the WLB + AFB exam by the physician (Table 1). There were three categories of visual grading: grade 1 for normal, grade 2 for abnormal (suggestive of inflammation, metaplasia, hyperplasia, and mild dysplasia), and grade 3 for suspicious (suggestive of moderate dysplasia, severe dysplasia, CIS and IC cancers) [4]. The total number of sites that had a WLB + AFB visual grade, a Raman spectrum, and a matching histopathology assessment from a biopsy was 193, of these 38 were positive for HGD and 24 were positive for IC/CIS based on histopathology (see Figure 5). The STEP-PLS diagnostic algorithms based on Raman spectra of these 193 sites has an area under the ROC curve of 0.85, very close to the value, 0.88 for the STEP-PLS algorithm generated from the full 280 tissue sites. When comparing the WLB + AFB visual grades to the corresponding histopathology, WLB + AFB grade 3 (suspicious) identified 27 out of 62 lesions that were moderate dysplasia or worse, representing a 44% sensitivity; the specificity was 68% because it correctly identified 89 of the 131 lesions that were mild dysplasia or lower pathology grade according to Figure 5. Alternatively visual grades 2 and 3 combined (suspicious + abnormal) identified 59 out of 62 lesions that were moderate dysplasia or worse, a sensitivity of 95%; the specificity decreased to 13% because only 17 of the 131 benign lesions/normal sites were correctly identified (Figure 5). The posterior probabilities generated from the STEP-PLS Raman algorithm identified 24 more HGD and malignant lesions (17 HGD, 1 CIS, and 6 IC) than the WLB + AFB grade 3 while keeping the same specificity level of 68%. Thus the relative sensitivity improvement of WLB + AFB+Raman versus WLB + AFB grade 3 was \((27 + 24)/27 = 1.89\). Alternatively if we keep the same sensitivity of 95% provided by the

![Figure 5 The histopathology distribution of bronchoscopy (WLB + AFB) visual grading of lung lesions and normal tissue sites. The number in the parenthesis is the total number of cases within a particular visual grading subcategory/histopathology subcategory combination.](image)
WLB + AFB grades 2+3 combined, the same Raman algorithm (STEP-PLS) were able to correctly identify 48 more benign lesions/normal sites (18 normal, 1 inflamed, 13 metaplasia/hyperplasia, and 16 mild dysplasia) than WLB + AFB grades 2 + 3 combined. Thus the relative specificity improvement of WLB + AFB + Raman versus WLB + AFB grades 2 + 3 was (17 + 48)/17 = 3.82.

5. Discussion

The results from multivariate statistical analysis demonstrated that HGD and malignant lung lesions can be differentiated from benign lesions and normal tissue using point laser Raman spectroscopy. We have previously shown that this was possible using a model algorithm based on linear discrimination analysis (LDA) of spectra processed using a 2nd order derivative to remove the background fluorescence [20]. The findings in the current study were obtained for a much larger sample size, using a system with alternative components to reduce system noise, and focus solely on the pure Raman spectra, rather than their noisier 2nd order derivatives.

The spectral range from 2775 cm\(^{-1}\) to 3040 cm\(^{-1}\) was chosen as the optimal wavenumber region for analyses as it contained the only significant Raman emissions in the range from 2050 cm\(^{-1}\) to 3100 cm\(^{-1}\) which was measured. These emissions were assigned to various CH stretching modes, although a small peak at 2790 cm\(^{-1}\), which helped with the discrimination between groups, was unidentified. The optimal upper bound cut off was determined to be 3040 cm\(^{-1}\) as the signal beyond this wavenumber had significantly more noise. Raman emissions due to water molecule stretching modes that occur for wavenumbers from 3200–3500 cm\(^{-1}\) were outside the range of our system.

There were some specific characteristics in the main Raman peaks that correlated with histopathology grades. Invasive carcinoma, for example, had spectral shapes that consistently demonstrated a reduced intensity of the CH\(_2\) symmetric stretch emissions at 2850 cm\(^{-1}\) and increased intensity at 2940 cm\(^{-1}\) from CH\(_3\) asymmetric stretching modes. Movasagi et al. suggested that the 2850 cm\(^{-1}\) was a good indicator for the change in the amount of lipid in the samples [36]. The 2940 cm\(^{-1}\) peak has been associated with both proteins and nucleic acids [20, 36]. A previous study that used Raman spectroscopy on ex-vivo lung tissue samples also found a reduction in the lipid content for malignant tumor tissue [41]. This was attributed to the decrease of phospholipids found in the cancerous tissue when compared to normal controls. The same study also found that there was an increase in certain amino acids while others decreased, [41] and many of these are known to have strong Raman emissions in the 2775 cm\(^{-1}\) to 3040 cm\(^{-1}\) range [20]. Thus changes in the abundance of these amino acids may have contributed to the changes seen in the Raman spectra measured in this study.

It was also clear that the average spectrum of inflamed tissue was significantly different when compared to those of other histopathologies. Although there were only a small number of spectra from inflamed tissue, all six cases came from six different patients, indicating that the abnormalities in the inflammation spectra were most likely not due to chance. Inflammation has been suggested to be a risk factor in certain cancers [42], particularly chronic inflammation. All six biopsies from the inflammation sites contained both chronic and acute inflammation histopathology, meaning that the Raman spectra could not be separated into the two subgroups. Due to this co-pathology, it remains unknown if the difference in the Raman readings from inflamed tissue was due to the chronic or acute diagnosis. Nevertheless the unique Raman signatures of inflammation tissues support the idea that Raman spectroscopy can help to improve the diagnostic specificity of AFB because inflammation is a significant cause of false positives in AFB lung cancer localization [13].

The change in intensity of individual Raman peaks in spectra from lung tissues with different histopathology were found to be insufficient for reliably predicting the histopathology of a random lung site. Multivariate classification techniques fared much better in predicting the histopathology of a random lung site. It was found that the ROC curves generated by PLS and PC-GDA analyses on the full (2775 cm\(^{-1}\) to 3040 cm\(^{-1}\)) spectral range had similar AUC of 0.83 (Table 4), and the AUC increased to 0.88 with STEP waveband selection. Two other waveband selection methods (LASSO or GA) returned similar AUC values, indicating the discrimination results were reliable. The sensitivities and specificities generated from these ROC curves were high (Table 5). At 90% sensitivity, 65% specificity is achieved by the STEP-PLS method in localizing HGD and malignant lesions. In comparison the AFB clinical trial in Ref. [4] that led to its FDA approval had a similar specificity of 66%, but the sensitivity is 67% only.

A trend for the spectra as a whole was apparent as shown in Figure 3C. A relatively large change in the posterior probability occurs between the mild dysplasia and the moderate dysplasia, indicated by the differences between the groups in the box plot. This high probability trend continues for Raman spectra from malignant tissue sites. All four multivariate techniques showed this trend, and the algorithm generated by using only the malignant cases and normal tissue spectra also showed the same...
trend (Figure 4C). These would appear to indicate that there was a significant change in the spectra from tissue with HGD compared to those from benign or normal tissue, and this change becomes more pronounced for malignant tissue. This is consistent with increasing severity of bio-molecular changes that accompany HGD and malignancy transformations respectively. Currently it is known that not every HGD will progress into a malignant state [11, 43], and those which progress are not able to be visually determined by AFB or WLB endoscopy from those which do not. Understanding the biochemical signatures of those lesions which spontaneously regress versus those which do not, would provide valuable clinical information. Early detection, especially detection of lesions which would be known to become invasive could improve therapy. The trends in Figure 3C and Figure 4C suggest a hypothesis that the posterior probability values calculated from Raman spectrum may be an indicator of the likelihood of a HGD lesion progressing to malignancy. The posterior probability values may also be used to predict the prognosis of invasive cancers. These hypotheses are worth to be tested in future studies.

Although the adjunct use of AFB to WLB has significantly increased the detection of HGD compared to WLB alone, there remains the inherent problem of poor specificities [10]. The specificity of the AFB visual grading used in this study ranged from 13% for grades 2 + 3 to 68% for grade 3 only. The corresponding sensitivities swung drastically from 95% to 44% respectively. Although the latter was a little lower than the values obtained in some AFB studies for similar specificities, it was not inconsistent with the high variability in sensitivities for high specificities as shown in the meta-analyses of Sun et al. [10]. This variability highlights how difficult it is to consistently obtain high sensitivities with good specificities. The adjunct use of Raman spectroscopy made a significant improvement to the specificity of detecting lesions in this study with HGD or worse. If high specificity is required, it was shown that for keeping the 68% specificity achieved by WLB + AFB visual grade 3, Raman identified 89% more true positives (51 sites compared to 27 respectively), representing a 1.89 times improvement in sensitivity. Alternatively if high sensitivity is required, it was shown that for keeping 95% sensitivity achieved by WLB + AFB visual grades 2 + 3 combined, Raman reduced the false positives by 42% (66 sites compared to 114 respectively), representing a 3.82 times improvement in specificity.

This study showed that the adjunct use of Raman spectroscopy improves the specificity of detecting HGD or malignant lung lesions of the main airways compared to AFB visual grading. No loss in sensitivity by Raman diagnosis occurred relative to WLB + AFB visual grades 2 and 3 combined; but the specificity increased significantly. Other methods have been tried to improve specificity, but with limited success. The fluorescence R/G ratio method looks particular promising but so far only one study showed any significant benefit [44], but its performance at high sensitivity settings are still inferior compared to our Raman results: at 95% sensitivity Raman has a better specificity of 49%, compared to the 32% for the fluorescence R/G ratio method; at 90% sensitivity Raman has a better specificity of 65%, compared to the 53% for the fluorescence R/G ratio. These results lend support to the concept that Raman spectroscopy provides the physician with objective secondary information compared to the more subjective visual appearance of lesions using WLB/AFB examination. This will reduce the number of false positive biopsies, procedural time, and healthcare costs. Most importantly, maintaining high sensitivity (≥90%) is critical for accurate diagnosis and management.

Raman scattering within the main lung airways can be measured within 1 second and used to improve the localization of lung cancer and precancerous lesions. We envision that an algorithm derived from a database of Raman spectra would be able to classify a lesion in less than half a second, making this approach feasible for real-time lung cancer localization. Different from subjective interpretation of WLB/AFB images, the Raman algorithm represents an automatic, objective diagnosis based on quantitative Raman spectral analysis. Point Raman measurement on lesions identified by WLB + AFB visual grading 2 and 3 combined is a promising new clinical method for real-time localization of lung cancer/precancerous lesions.

6. Conclusions

In conclusion, a single centre clinical investigation of the adjunct use of real time Raman spectroscopy to the standard WLB and AFB for in vivo lung cancer localization of the central airways was conducted. In vivo real-time point laser Raman spectroscopy was performed on 280 lung lesions and normal tissue sites with a measurement time of 1 second per spectrum. Using multivariate techniques and waveband selection methods on the Raman spectra, it was shown that HGD and malignant lung lesions can be differentiated from benign lung lesions and normal lung tissues with high sensitivity (90%) and good specificity (65%). Compared to WLB + AFB visual grade 3 based diagnosis, Raman + WLB + AFB improved the sensitivity of localizing HGD and malignant lesions by 1.89 times, while compared to WLB + AFB visual grade 2 + 3 combined diagnosis, Raman + WLB + AFB improved the specificity of
localizing HGD and malignant lesions by 3.82 times. Different from subjective interpretation of WLB/AFB images, the Raman algorithm could potentially become an automatic and objective diagnosis method based on quantitative spectral analysis. Further multi-center clinical trials are warranted to fully test the potential of this technology for improving lung cancer and precancerous lesion localization.

**Supporting Information**

Additional supporting information can be found in the online version of this article at the publisher’s website.

**Acknowledgements** This work was supported by the Canadian Institutes of Health Research (CIHR grant #PPP-79109 and MOP-85011) and the Canadian Cancer Society Research Institute (Grant # 20352). The authors would like to thank Myles McKinnon, Jennifer Campbell, and Dr. Keishi Ohtani for endoscopy procedure support, and Dr. Martial Guillaud for providing statistical expertise for the Raman data analysis.

**References**


